plausible possibility of pointing at clusters of place holders that may either be identified with one new topic relation or several new topic relations to be dealt with in one act of innovation.

The other restriction is that A is bound to model a new topic relation (Stage (a) in Section 3.2) in one modelling facility. The restriction is again inessential. The most damaging consequence of the constraint is that A is discouraged from inventing topic relations (perhaps the most important kind) that are analogies between different parts of a heterogeneous field. Such inventions are possible, by dint of setting up special L^1 descriptors and/or L^0 primitives in an (otherwise homogeneous) modelling facility and this ritual can be performed in the system as it stands. But it is unnecessarily tedious and is circumvented in the system under construction, where it will be possible to make several independent models on the same occasion, n, and to execute them with or without interaction.

If a source's thesis about a subject matter satisfies the consistency and cyclicity requirements, then an entailment structure can be derived by the methods discussed in the last two chapters. Given this structure there are simple algorithms for generating all the multiple choice or list questions (the $\underline{\text{PQuest}^0}$s) that satisfy any given specification; in particular such questions are defined for each topic relation R_i in the conversational domain R, from which the entailment structure is derived. For each R_i it is also possible to issue a base command to bring about R_i in some or any way; it is part and parcel of the construction procedure that the source can obey such a command, usually by making a model which, on execution, brings about R_i. So far as the source is concerned, the model can be made in the abstract; as an explanation of R_i. But, unless we are prepared to interpret a student's explanations (by analysis of L^0 expressions) the source must furnish a standard modelling facility or, more often, several of them appropriate to different parts of the conversational domain. In fact, the act of describing the models built in a modelling facility is (usually, in our experience), the easiest way of interpreting explanations, though other methods of interpretation are possible. The non-verbal explanations of R_i that are counted as legal according to the source's thesis, constitute some (and without special qualification all) of the models that can be built in the facility proper to R_i and that, if executed, bring about, maintain, or satisfy R_i.

This chapter is concerned with modelling facilities. As a first step (Section 1), it is useful to consider a slightly more general category of environment, namely *laboratories;* (meaning workshops, construction shops, studios, demonstration rooms in addition to special scientific laboratories). Laboratories are familiar and an analysis of the activities that go on in them illuminates the nature of modelling as well as casting some scattered light on the modelling that might go on in facilities slightly more liberal than those that currently exist. It is useful to suppose, throughout this part of the chapter, that any operation carried out in a laboratory is sensed electrically (like the plugging up and meter setting operation in STATLAB). Nowadays, this claim seems quite reasonable and if the reader is disturbed by it I am quite prepared to justify the practical possibility of electrically sensing any operation of interest.

The remaining sections are also written under this assumption and treat of various kinds of laboratory like modelling facilities appropriate in connection with different subject matters. Unless the contrary is stated, it is assumed that the modelling facilities are single clocked (in the sense of Fig. 1 in Chapter 5) in order to retain a structure/function demarcation i.e. to separate model building/construction from model execution. Also, the class of models that may be built is limited to the class of finite automata or finite state machines.

Section 2 is devoted to modelling operations that are reasonably construed as problem solving operations; for example, programme writing, constrained by certain limits and goals. In Section 3, it is argued that the same basic paradigm adumbrates situations that are commonsensically regarded as *games* and/or as *control* situations (under the formulation employed, problem solving and control are special cases of "constructing game strategies" which are, in this context, "non-verbal explanations of play").

Section 4 extends the discussion to *metagames* and Section 5 digresses slightly to highlight an interpretation of metagames as situations involving interpersonal interaction. Due to the one clock constraint, it is impossible (except in trivial cases) to elicit models that explain, for example, policies, or motives. In particular the (existing) facilities cited in Section 5, are limited to posing PQuests and eliciting replies rather than posing EQuests that elicit explanations. Section 6 is a very brief account of the capabilities likely to be added if this constraint is lifted, as it can be.

1. Laboratories

In physics, chemistry or biology explanations are elicited and demonstrations are given by modelling operations carried out in laboratories: The careful observation of a "practical" class is usually agreed to furnish a great deal of information about the participating student's intellectual capability in terms of how he explains events. It is worth remarking upon the various roles played by laboratory work for precisely the same remarks are applicable to all of the simulations, modelling facilities, etc. under discussion.

1.1. *Techniques.* At the most primitive level, a student learns techniques; for example, how to handle equipment or make accurate measurements. Using a simulator such as STATLAB this aspect of laboratory work is minimised by design, quite deliberately. In full scale laboratories it might be minimised with advantage, also. It seldom happens that technical expertise is an essential accomplishment. When it is, a conventional laboratory is about the worst place to acquire the requisite skills. The diversity of real fabric tends to obscure clearcut operations (handling materials, the minutiae of measurement, etc) which are defined as specific procedures; best and least painfully inculcated by formal condition charts or by training algorithms which prescribe and monitor the handling of tangible equipment.

1.2. *Quasi experiments.* Next, laboratory courses contain a series of preprogrammed "experiments" the outcome of which is known, supposing the "experiment" is properly conducted. Hence, the actual result or outcome is used, indirectly, to judge how well the student brought about the commanded relation (for example, identification of an "unknown" compound or a determination of Young's modulus).

There is no objection whatever to this aspect of laboratory work, apart from the use of the word "experiment" and the more or less explicit idea that discovery learning is fostered under these circumstances. In fact the criteria of veridiciality are standard and very abstract (nothing new is being discovered and the student generally knows it. If by lucky happenstance, he did make a discovery the aberrant result would not be credited).

Really, the preprogrammed quasi-experiment is supposed to represent a world having text book rules; and a student's performance is judged on that assumption. It was once true that the easiest way (often, the only practical way) to create a text book world was to carry out physical operations in the same domain; for example, chemical experiments in chemistry, or biological experiments in biology. Nowadays, this is rarely the case. It is usually better to simulate the text book world by a piece of electronics or a computer program; two examples are provided by STATLAB and by a "fruit fly breeding" quasi-experiment (in genetics) very beautifully programmed on PLATO (the student calls for examples of flies and sees them; he specifies certain matings of pairs; the program displays the text-book world offspring).

A very fundamental and important point is implied by this statement; and often goes unnoticed because the mere convenience of a simulation (clear, uncluttered presentation, and so on) is taken to be its principal merit. Undoubtedly, the presentation does matter. But the real distinction lies in the fact that quasi experiments can be simulated, usually in a rather simple minded fashion, just because nothing whatever is lost. The student is meant to learn about a text-book world governed by the "laws of physics" or the "law of genetics" (not about reality, except, of course, reality as it is pictured by the physical or genetic "laws"). Even if the student works in a real laboratory, every effort is made to ensure that he learns these "laws" and neither flirts with the unforseen nor exhibits unusual modes of explanation.

Real phenomena (whether concrete or abstract) have an indefinite number of explanation types that are acceptable though they are also open to an indefinite number of trivial explanations and an indefinite number of essays in explanation that are unacceptable because they are mistaken. For a text book phenomenon there is one class of equivalent explanations deemed "correct" and all members of the "correct" class are recognisable[21].

So, also, some problems have an indefinite number of acceptable solution types. But others do not. If an experiment is a quasi experiment then the phenomena to be explained are presented as posing problems that have very limited, well known and thus accurately recognisable solutions.

1.3. *The Hinterland.* Some experiments lie on the borderline between quasi experiments (really problem solving situations) and genuine research. The instructor names an imperfectly specified goal and the student casts around to discover an insightful way of achieving it. His progress is observed; innovation and (genuine) discovery learning is encouraged.

In this connection it is cogently argued that the variety of real world fabric is a crucial determinant of insight, absent from the majority of simulations. The claim in favour of "variety" is accepted without qualification; but, since the case under

[21] This does not mean that there is one and only one correct explanation, like the answer to a multiple choice question. The one correct class may have an infinite number of members and commonly it has many members.

consideration is borderline, there is some doubt about what kind of "variety" is involved and how it is used.

Two possibilities can be mechanically accommodated; one cannot be (as a matter of fact, this remaining possibility cannot be "accommodated" in any manner, mechanical or otherwise). Of the two tractable cases only technical and generally trivial limitations stand in the way of less arid simulations.

(a) The "variety" in question is a variety of instances or examples which cover the field of whatever relation (some R_i) is brought about if the goal is achieved or the problem is solved. There is no more difficulty in generating the requisite exemplars (specimens of fruit flies, perhaps) by electro-mechanical methods than there is in collecting the exemplars for inspection; often, the simulated generation process is a great deal less difficult. Here, the variety is used to exhibit the relation to be learned (on the assumption that the relation is known by the teacher). Since statistical relations, at any rate in a stationary universe, are perfectly knowable, it follows that provision of a random input tape to STATLAB or the random mutation and combination of fruit fly gene patterns to produce statistically legal offspring, is an operation that comes into the same category as the generation of exemplars; in the sense that it has the same cognitive impact and function in teaching (i.e. a statistical law is treated as an ensemble of deterministic laws).

(b) The other (tractable) case is somewhat different. Without prejudice to whether or not the relation (R_i)to be learned is exhibited by running over the exemplars in its field or (for a statistical law) over the deterministic laws it adumbrates, the student is encouraged to seek novel explanations for R_i. Here "variety" means "novelty" which is produced by the student, not by features immanent in the subject matter.

Notice, first, that non trivial modelling facilities usually can interpret several forms of explanation, and are made to interpret as many classes of explanation (perhaps with infinite membership) as the designer specifies. So far as novelty is concerned, the important point is that there should also be criteria saying that some classes of explanations are prosaic or mundane or obvious. If such criteria exist, the student who submits an obvious explanation can be told to seek a better or more subtle one and the act of rejecting a "correct but pedestrian" essay (whilst confirming its correctness) does encourage novelty, for good reasons. In fact, this expedient is employed by one of the "evolutionary" CASTE heuristics which is described later, when the "good reasons" are also presented. So, if criteria for "prosaic" and "obvious" are specified for several interpretable classes of explanation, then the exclusion of unimaginative explanations is readily mechanised and the facilities required can be embodied in a modelling facility. Conversely, if these criteria are not given, then modelling work is restricted to pseudo experiments; regardless of whether it is done on a simulator or in a room full of apparatus manned by human instructors.

Finally, consider the one intractable case. Here, the "variety" is an indication that an indefinitely large number of explanation types may be deemed acceptable. The notion of "acceptability" is either inherently human (depending upon consensus of opinion, the explanation type currently in vogue, societal norms) or it is the variety that remains in nature (the genuine richness of fabric) even if a very well defined procedure like scientific verification is brought to bear upon putative explanations. In this case, explanations remain forever as hypotheses and the student may, at any stage, become the subject matter expert, the source, or a professional scientific investigator.

1.4. *Genuine experiments that involve discoveries.* A scientific investigator (equally a student in that role) is cabined either by a real laboratory or a simulation. Only certain experiments can be done with the equipment deployed in different combinations. In precisely the same way, any observer is confined within the limits of his senses. A solipsist, alone, is free to invent in some more general sense. These restrictions upon method, apparatus and approach have their uses as well, of course; it is hard to see how any coherent experimentation could go on without them, for they constitute a frame of reference (biology, or physics, or genetics) within which order can be discerned in the flux of existence. The main reason for bringing the matter to attention is to point out that a laboratory is not an unmixed blessing and that whenever we talk about genuine experimentation or innovation (as I did in the last subsection) we are referring, in fact, to an activity that is a compromise between unfettered creative volition (the solipsist making his mental construct and projecting it outside) and the boundaries imposed by a "frame of reference".

The obtrusiveness of these boundaries is, in a peculiar way, dependent upon context. It would be fruitless, at this juncture, to

take up the ontological issue of whether (for example) a physicist needs a cloud chamber to be a physicist or whether he could play at physics in a different arena. The immediate point is that a frame of reference is given and any laboratory amplifies the restrictions, thus imposed, by various constraints that are prerequisites for capturing the essence of material or process. In this respect, the machinery in which a modelling facility is embodied (usually a computer together with a programming language and compiler routine) is less limiting than most real life laboratories and in some contexts is peculiarly unrestrictive. For example, it can (appropriately organised) mediate experiments with virtually unlimited data structures; Papert and others have often noted that first and foremost, a computer is a mathematics laboratory for genuinely innovative experiments[22].

1.5. *Rules and roles in laboratories.* A respondent, placed in a laboratory, is faced with rules. Within the confines of the laboratory he acts out different roles; student or, investigator or, subject matter expert.

1.5.1. *Rules.* The first kind of rule is an esoteric language. The jargon of inorganic analysis is, perhaps, the prime example of narrowness in this respect; but some of the humanities rival it quite closely. Any science from physiology to anthropology has a jargon of one kind or another and other-than-jargon statements are open to misinterpretation.

[22] Though people frequently pay lip service to this principle it is rarely applied in practice. There are notable exceptions, Wallace Feurtzig, at Bolt Beranek and Newmann; Peter Jensen, at Georgia Tech.; Seymour Papert, at MIT and James Thomas, at Brunel. But, as a rule, computing facilities are used to mimic or try out ideas that exist (algorithms, process schemes, abstract models). This circumstance is partially attributable to lack of resources and to the difficulty (even using an on-line system) of communicating with the machine in a humanly intelligeable manner. But there is also a confusion between "testing a model" and the inventive act of "building a model" and this, it seems, is the real obstacle. For example, the computer is rarely used as an aid to invention in operational research, (however much the practitioners say it is) just because invention is seen as going on in the brain only; not really in the medium of a computing system. The managers who are accused of mis-using operational research models by mistaking them for facts are guilty of a reversed version of the same fallacy; they conceive the computer program that represents stock holding (for instance) as the actual inventory system and its predicted future, just because of an ingrained idea that a representation outside a brain must be a fact.

Next, there are rules of deportment concerned with priority and obescience (to ideas, rather than people); the courtesies of reporting and publication, together with values that are imposed upon experimental goals (the creditably, originality, etc., of an investigation).

The next kind of rule is an assembly rule; there are certain pieces of apparatus and certain reagents. For physical and conventional reasons, these can only be assembled in certain ways. Paradigmatic configurations are specified by "set" quasi experiments. But, in general, it is possible to construct new apparatus from special configurations of the existing building blocks, and possible to order in new apparatus from outside suppliers.

Finally, there are rules of interpretation. Though results are expressed in the esoteric language and are subject to the rules of deportment, further canons are applied (namely the interpretation rules), in judging whether or not a modelling operation, described as an explanation, is deemed acceptable to the laboratory "establishment".

1.5.2. *Roles.* The respondent, placed in a laboratory and obeying these rules, is classified according to the intentions he entertains or is meant to entertain.

If he is a student under instruction, then he is required to keep the rules (all of them) and is penalised for not doing so. His intention is to obey (learning how to achieve obedience when necessary).

If the respondent is a professional investigator then he is encouraged to break some but not all of the rules. His intention (phrased in the vernacular of psycholinguistics) is to produce an "ungrammatical but interpretable utterance". The comment (in section 1.4) that constraint is one prerequisite for progress states the fact that it is impossible to break a rule unless the rule exists.

1.5.3. *Discoveries.* Several "kinds" of "discovery" are commonly distinguished according to the dominant type of infringement.

Axiomatic innovations change the esoteric language but leave the other rules unchanged.

Theoretical discoveries break the rules of interpretation (with or without any axiomatic displacement) but leave the other rules unchanged apert from an extension in scope.

Empirical discoveries are attained by breaking an assembly rule and leaving the other rules unchanged though their scope is usually extended.

Mixtures of these components are important. For example "large and fashionable" discoveries are generally associated with several types of infringements, which lead to a result judged "plausible" in terms of some deportment rules.

1.5.4. *Special status of deportment.* The rules of deportment have a special status insofar as they cannot be broken and revised by an investigator at his own whim. They are periodically reviewed and revised by a responsible body (meetings of scientists, scholars, administrators). Any discovery is liable to infringe the status quo; but it does so mildly and the mores of science are constructed to tolerate this much deviance. The face saving deportment rule is the liberal dogma that all models, or the explanations that stem from them, remain as hypotheses open to denial. The deportment rules are changed when they are mildly infringed by several conjointly consistent discoveries, usually quite small discoveries in their own right. Hence, these rules evolve, and are admitted to do so. But they evolve gradually.

An outstanding "large" discovery infringes many deportment rules; for example, it may be impossible to report it in journal format and it may not seem relevant to the "mainstream" of thinking. It is often tentatively accepted, nevertheless (and looked at askance for a while). But, because the deportment rules are known to evolve (science is blessed with self-consciousness), the currently offending discovery may be seen to "possibly fit" a future structure.

1.5.5. *The position of a subject matter expert.* A source (subject matter expert) either has the calibre of an investigator, or a meta-investigator, or both.

As a preliminary, notice that the rules of esoteric language, of deportment, of assembly and of interpretation are themselves relations, with the peculiar property that any topic relation in the sense of the last chapter (and in common sense also) involves all of them; so does its class of descriptors.

Various clusters of topic relations count as theories credited to a source (for example, "Jameson's theory of biology"), independently of their truth/falsity or whatever, insofar as the source (Jameson) can satisfy the technical requirements of consistency and cyclicity (last chapter), insofar as he can describe the result and (the task we are now approaching) provided he can say how to model or bring about each topic relation.

Very many possible theories exist and though the source (Jameson) typically addresses theory construction as a creative task, he might be imaged as making a selection from this very large set.

Equipped with that preamble, it is possible to distinguish between sources.

1.5.5.1. If the source is an investigator, he selects his current theory. This may be very difficult, for what he calls a "current theory" is often beset by gaps that disallow it as a theory in the "technical" sense.

1.5.5.2. If the source is not an investigator, then he forms an amalgam of topic relations, mooted by other authorities, which does satisfy the technical requirements. Further, he uses canons and criteria and performs experiments in order to ascertain (in some sense) a "best" theory or a "best" class of theories, to which he appends his name (not, like the investigator, to mark "his discoveries", but "his images of the present status quo").

A source who does this job fully and honestly, functions as an innovative investigator. He differs from the investigator, qua purveyor of idiosyncratic expertise, because he experiments on a domain of knowables instantiated by laboratory tasks, rather than experimenting in a laboratory.

1.5.6. References to a "student", rather than a "student under instruction" pick out a respondent somewhere on the continuum "student under instruction" or "source" or "professional investigator". Though we cannot deal with the liberal extremity very well at the moment, attention is certainly not confined to the "student under instruction".

1.6. *The generality of laboratory facilities.* It is possible to construct laboratories and the corresponding modelling facilities for any subject matter whatsoever; for example, for history or economics.

Curiously enough, quite ordinary facilities that count as laboratories according to the present definitions, are often overlooked, though the definitions themselves are neither odd nor outlandish. One (generally unnoticed) laboratory a "dance laboratory", is a stage together with the esoteric language afforded by a rich choreographic notation (whilst very differently conceived the notation schemes due to Eshkoll, Benesh, and Laban are all competent in this respect though Eshkoll's, only, is specifically adapted to computer manipulation). Similar comments apply to art/architecture or design studios and the private languages employed when people use them. Some of the private languages are bundles of idiom and dialect, wrought by tradition from the natural tongue (by a history of master/apprentice rapport, for example). Others are deliberately engineered extensions of natural language; for example, in their study at St. Martins School of Art, Laurie Thomas and his associates use repertory grid techniques, augmented by cluster analysis programmes. These programmes are developed, with student participation, during the study, from a few primitive originals. Very similar methods are used in the theatre by Joan Littlewood and have been widely diluted and plagiarised (the name of her own company "Theatre Workshop" indicates that she, at any rate, has regarded the theatre as a laboratory since the early 1950s). As a final instance, sensitive driving instructors use a motor car and the surrounding traffic as a laboratory; during those phases of training when procedures are demonstrated, and plans, tactics, and intentions are discussed with the learner, after or during a drive.

1.7. *Homogeneity of laboratories.* Provided there are assembly rules and interpretation rules (Section 1.5) there is no reason to reject heterogeneous laboratories in which models are constructed over diverse fields. In fact, heterogeneity is quite commonplace, even if the laboratory is designed for well structured subject matters (physics, for example, is split into disparate compartments for tutorial purposes; hence, the physical laboratory contains an optical bench for "light" and an assortment of resistance boxes, galvanometers and so on, brought out as a prelude to exercises in "electricity"). Homogeneity is rare, except in some very restricted branches of mathematics.

That this may be so, follows from the last chapter where it was noted that for R_i, R_j in a conversational domain R it is quite usual to find that the field of $D^0(R_i)$ is incomparable with the field of $D^0(R_j)$ apart from the entailment coupling expressed in $D^1(R)$.

We do, however, require that the models made in one compartment are built according to certain common principles (such as complementation and composition) and thus a modicum of communality between the assembly rules and the interpretation rules proper to each compartment in question. Formally, this requirement is satisfied by the existence of an entailment structure. But the formalision could express trivial correspondences and it is quite clear that effective learning over a heterogeneous field depends upon the existence of many strong analogies between the distinct compartments; for example, that principles learned in "light" can be applied, as analogies, to the study of "electricity".

1.8. *The veridiciality of laboratories.* It is quite important to notice that a laboratory could be counterfactual: for example, an Ames Room (or on a larger scale, the grotesqueries and phantasmagoria of Disneyland) furnish a creditable laboratory. So would the type of set up used by Michotte in his experiments upon apparent causality. The fact is, the canons for accepting a model built in a laboratory are satisfaction or "workability" (that any model must be executed) and agreement over explanations offered in the esoteric language (Section 1.5) which depends upon the rules of etiquette, deportment, and interpretation. A model (as stressed repeatedly in the last chapters) represents something teachable but the question of truth and falsity does not bear directly upon these properties.

1.9. *The laboratory as an extension of the user.* It was noted in Chapter 5 (Section 2.5.8) that one modelling facility or laboratory is the P Individual's biological processor ("his" M Individual); further, that it is sometimes useful to regard a facility such as STATLAB, or even a motor car, as an extension of this processor that in some ways enhances, in other ways restricts, its capabilities. Henceforward, we shall keep that interpretation in mind as generally salutory.

Certain corollaries follow, as a result of this orientation. Whatever is known to a participant (the relations he takes as properties and that are consequently the functional primitives) are no more nor less than relations he can bring about by doing (by way of model execution) in the processor (biological entity of mechanical modelling facility) in which this P Individual is executed. Thus, for example, I can walk, or press in plugs, or distinguish up from down, or one texture from another, or right from left.

1.10. *"Truth" in a generalised laboratory.* Return to the idea of "truth" and the discussion in Section 1.8.

There is a sense of "truth" to the participant P Individual in a conversation which runs as follows; the statement "That is true" means "I could do that and bring about that (true) relation, using the processor in which I am currently embodied and executed". Clearly this concept of "truth" is processor dependent; it may vary greatly from context to context and it is likely to change with the extent to which biological processors are (if at all) mechanically augmented or coupled into social aggregates for performing cooperative tasks. That tallies, well, with everyday experience. But "truth" (in this special though very useful sense) is really to do with satisfaction (by a procedure, of a relation) and cannot be directly equated to the "veridicial truth" of an external observer.

2. Some Laboratory Like Modelling Facilities

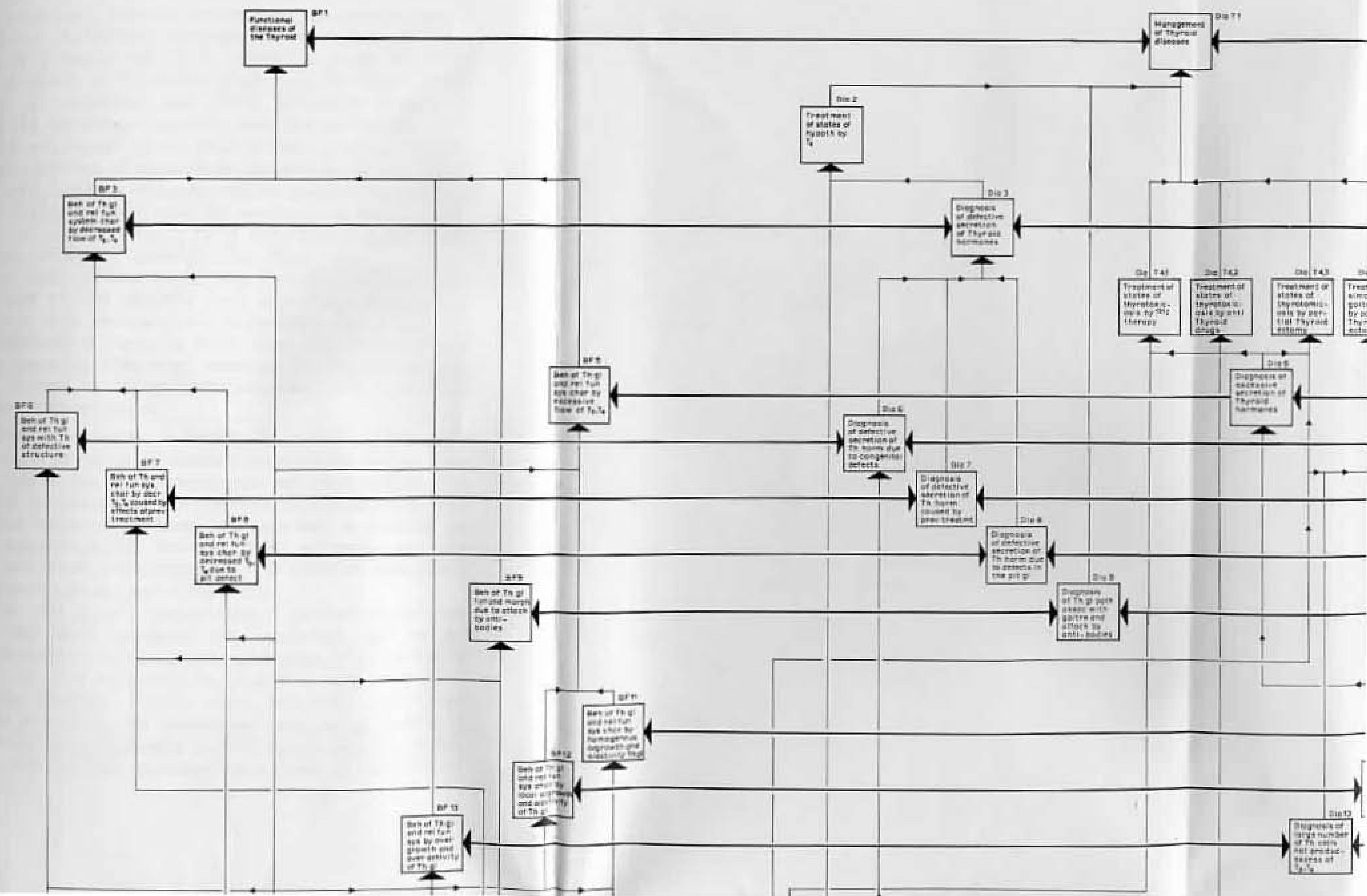
In Chapter 5 Section 1.7.4 we chose to constrain admissable models to the class of finite automata or of finite state machines with optional extension to locally metricised probabilistic machines (for which the state sets (internal, Z; input, U; output, V) are fixed, but the transition between any one internal state and another may be determined by a probabilistic transition rule, interpreted as an appropriately biassed random selection generator, reset to act as an independent random source at each execution step τ. In the same chapter we introduced a pair of clocking arrangements (Fig. 1 and Fig. 2 of Chapter 5) the modelling facility being serially clocked unless the contrary is stated. It follows that the operation of any modelling facility can be regarded as a special case of Icon 10, provided that the modelling facility forms part of CASTE.

2.1. *Retrospective comments on STATLAB as a modelling facility.* A modelling facility of this type is more elaborate than it need to be to accommodate the models generated under the main example of elementary probability theory. The two lower quadrants in STATLAB (Chapter 4) (the subfacilities for structural models and real life experiments) need only allow a student to construct graphs or mappings from the elements of a set onto a finite number of tokens for its subsets, the members of which are specified by the mappings in question. In the abstract quadrant, the set is an event set, from which elements can be excluded at the students discretion and the subset tokens (again included or excluded at the student's discretion) are simply markers for unions of elements called "composite events". Choice of a set of elements and a set of tokens determines a universe of modelling in which a model is the graph formed by "plugging up operations". In the "real life" subfacility the set is a set of "results" or independently observable occurrences taking place over a unit interval of execution, $\tau \to \tau + 1$, either at the students discretion or as determined by "nature"; the tokens represent subsets of "composite results". Once again, in this subfacility, a universe of modelling is selected by choosing as many "simple results" and as many subset tokens as required; a model is a graph formed by "plugging up" operations.

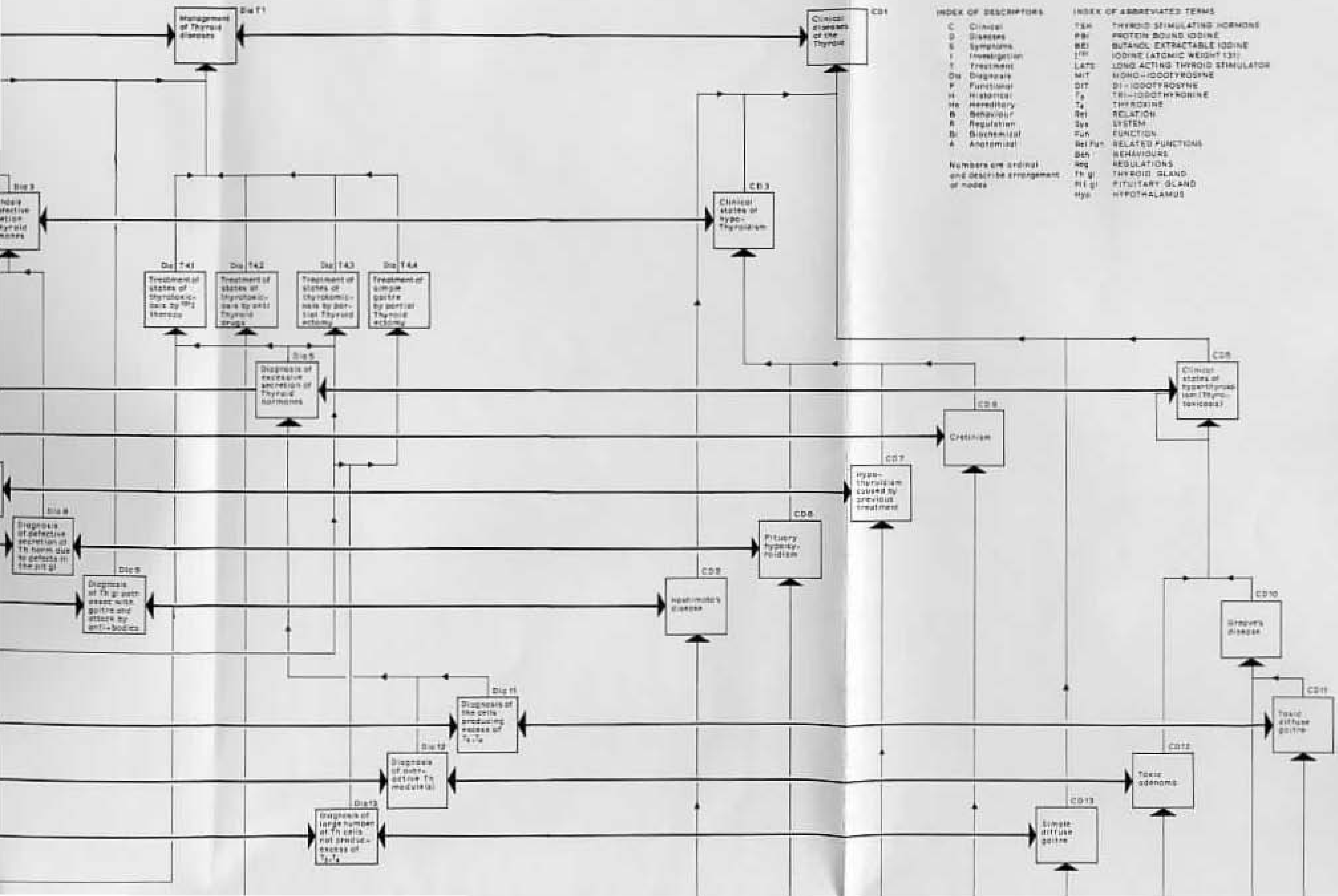
The other quadrants are not so simple. The Boolean operators (disjunction, conjunction, complement of any subset of results with respect to any other) in the upper left quadrant are trivially finite state machines i.e. they have a fixed and predetermined internal state and no other storage capability. The frequency counters are stores, of course, but being specialised as counters, they are degenerate. Hence, any admissable STATLAB model belongs to a very specialised and restricted class of finite state machines. Similar comments apply to the arithmetic operators "+" and "−" of the upper right quadrant but, in this case, the restriction is of a very different kind (although "+" is in register with union and "−" with relative complementation there is no arithmetic operation corresponding to intersection; hence any measure on a conjunctively specified subset must be expressed, in common terms, as a measure on the equivalent subset formed by disjunction and complementation). As a result, models built in the upper quadrants are analysed by much more complex routines, in the program of Appendix E. These routines match strings of several types of object, at various depths from the (one) outcome.

Mechanically speaking, the only probabilistic part of STATLAB is the random input tape (there are no random state transitions since the relevant finite state machines are reduced to fixed mappings). To justify the further degeneracy (that output states are merely counted, the count ratios being frequencies) the input tape of experimental results must represent a stationary "nature".

2.2. *Extensions of STATLAB.* The subject matter of "elementary probability theory" as it is exhibited in the



BF 1
Functional diseases of the Thyroid
Dis T 1
Management of Thyroid diseases
Dis 2
Treatment of states of hypoth by T_4
BF 3
Dis 3
Diagnosis of defective secretion of Thyroid hormones
Dis T 4.1
Dis T 4.2
Dis T 4.3
BF 5
Dis 5
Diagnosis of excessive secretion of Thyroid hormones
BF 6
Dis 6
BF 7
Dis 7
BF 8
Dis 8
BF 9
Dis 9
BF 11
BF 12
BF 13
Dis 13



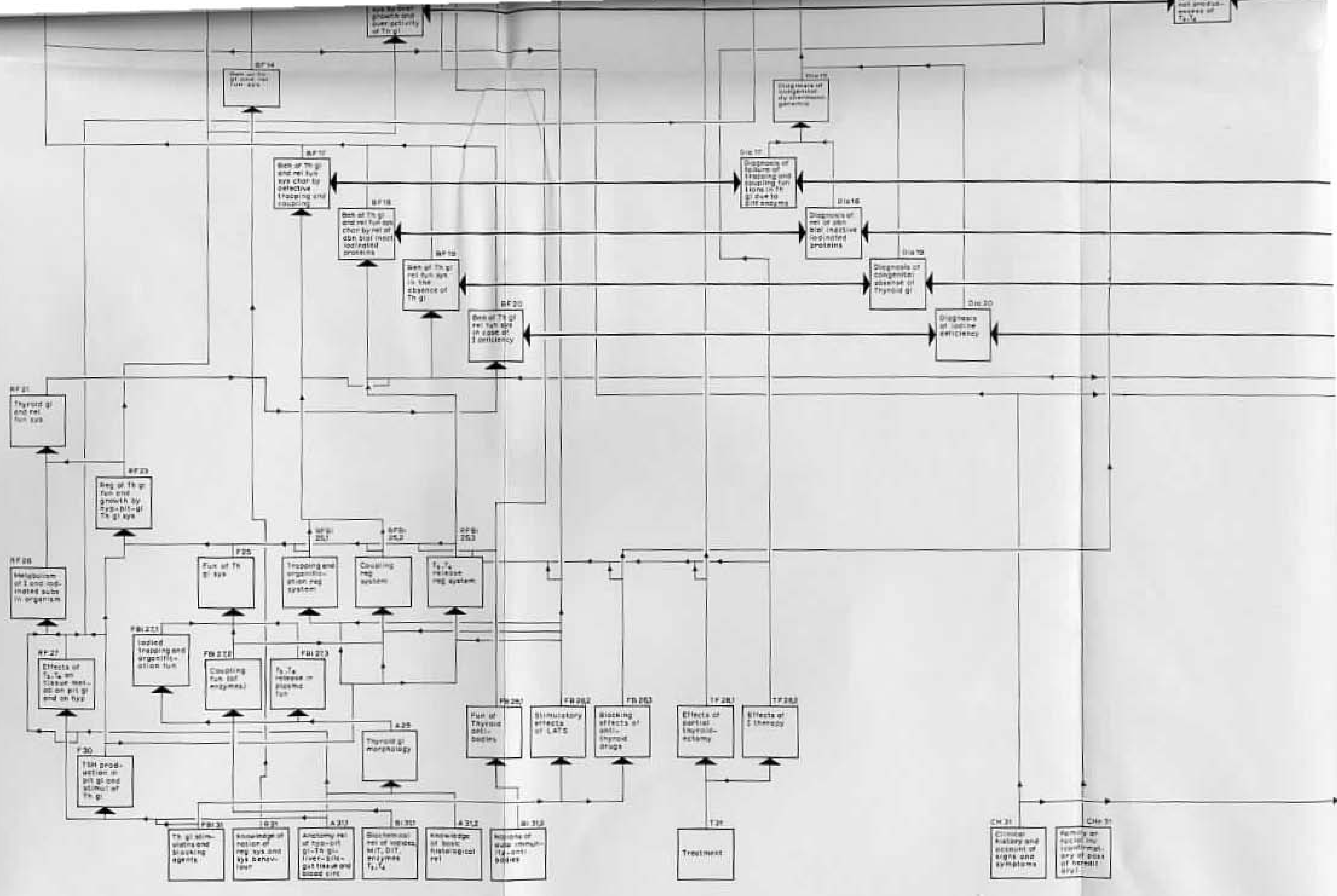
INDEX OF DESCRIPTORS

C	Clinical
D	Diseases
S	Symptoms
I	Investigation
T	Treatment
Dia	Diagnosis
F	Functional
H	Historical
He	Hereditary
B	Behaviour
R	Regulation
Bi	Biochemical
A	Anatomical

Numbers are ordinal and describe arrangement of nodes

INDEX OF ABBREVIATED TERMS

TSH	THYROID STIMULATING HORMONE
PBI	PROTEIN BOUND IODINE
BEI	BUTANOL EXTRACTABLE IODINE
I^{131}	IODINE (ATOMIC WEIGHT 131)
LATS	LONG ACTING THYROID STIMULATOR
MIT	MONO-IODOTYROSYNE
DIT	DI-IODOTYROSYNE
T_3	TRI-IODOTHYRONINE
T_4	THYROXINE
Rel	RELATION
Sys	SYSTEM
Fun	FUNCTION
Rel Fun	RELATED FUNCTIONS
Beh	BEHAVIOURS
Reg	REGULATIONS
Th gl	THYROID GLAND
Pit gl	PITUITARY GLAND
Hyp	HYPOTHALAMUS



BF 14
BF 17
Beh of Th gl and rel fun sys char by defective trapping and coupling
BF 18
Beh of Th gl and rel fun sys char by rel of dbn biol inact iodinated proteins
BF 19
Beh of Th gl rel fun sys in the absence of Th gl
BF 20
Beh of Th gl rel fun sys in case of I deficiency
RF 21
Thyroid gl and rel fun sys
RF 23
Reg of Th gl fun and growth by hyp-pit-gl Th gl sys
RF 26
Metabolism of I and iodinated subs in organism
RF 27
Effects of T_3,T_4 on tissue metab on pit gl and on hyp
F 25
Fun of Th gl sys
RFBI 25,1
Trapping and organification reg system
RFBI 25,2
Coupling reg system
RFBI 25,3
T_3,T_4 release reg system
FBI 27,1
Iodied trapping and organification fun
FBI 27,2
Coupling fun (of enzymes)
FBI 27,3
T_3,T_4 release in plasma fun
A 25
Thyroid gl morphology
F 30
TSH production in pit gl and stimul of Th gl
FBI 31
Th gl stimulators and blocking agents
Knowledge of notion of reg sys and sys behaviour
A 31,1
Anatomy rel of hyp-pit gl-Th gl-liver-bile-gut tissue and blood circ
BI 31,1
Biochemical rel of iodides, MIT, DIT, enzymes T_3,T_4
A 31,2
Knowledge of basic histological rel
BI 31,2
Notions of auto immunity-antibodies
FB 26,1
Fun of Thyroid antibodies
FB 26,2
Stimulatory effects of LATS
FB 26,3
Blocking effects of anti-thyroid drugs
TF 28,1
Effects of partial thyroidectomy
TF 28,2
Effects of I therapy
T 31
Treatment
CH 31
Clinical history and account of signs and symptoms
Dia 17
Diagnosis of failure of trapping and coupling functions in Th gl due to diff enzyms
Dia 18
Diagnosis of rel of dbn biol inactive iodinated proteins
Dia 19
Diagnosis of congenital absence of Thyroid gl
Dia 20
Diagnosis of iodine deficiency

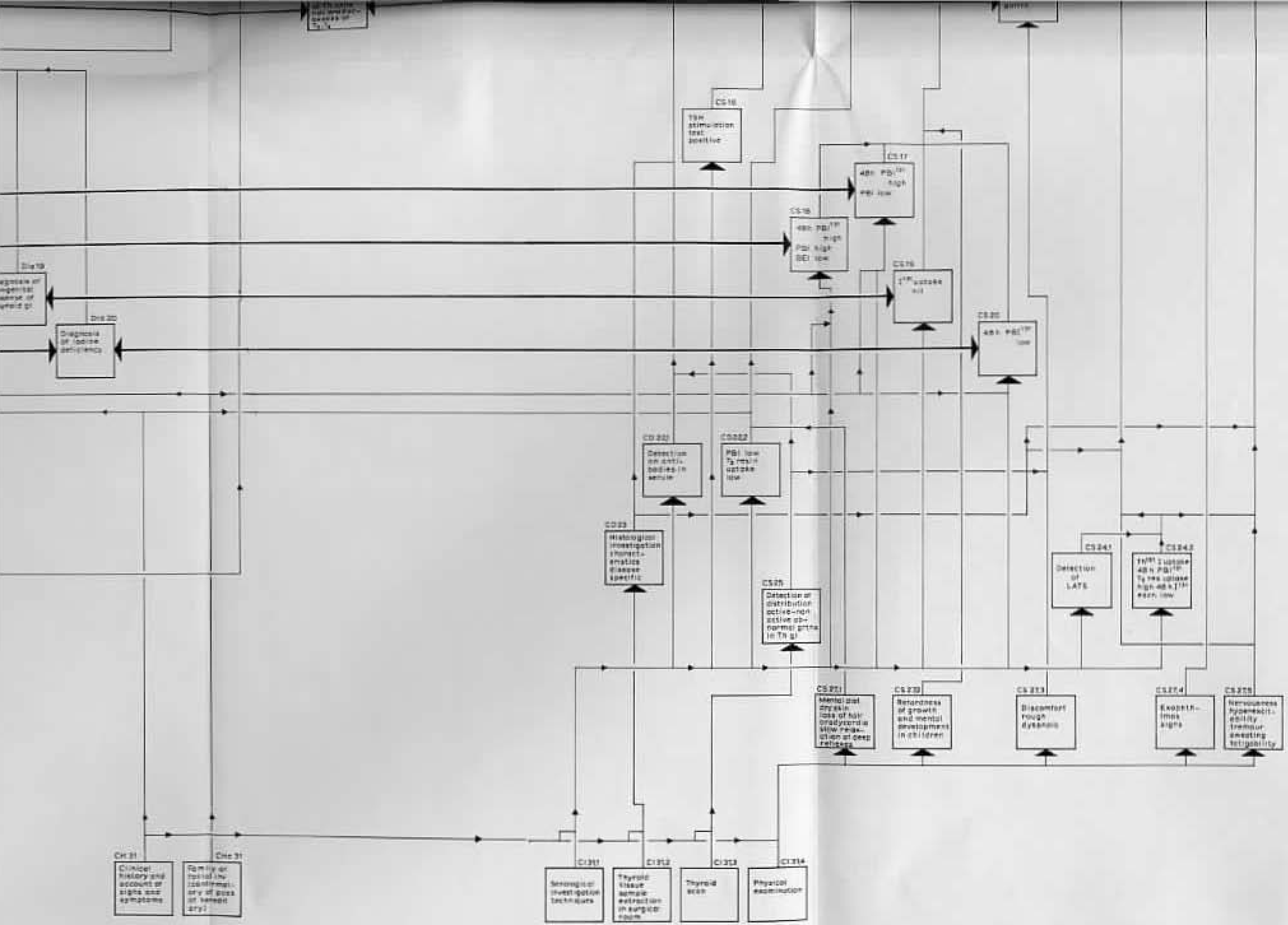


Plate 11. An entailment structure for a thesis on diseases or defects of the thyroid gland and functionally associated regulatory systems.

entailment structure of the main example was culled from a much larger entailment mesh (Plate 10, Chapter 7) described under the head "Chebyshev inequality (of variances) as an explanation for long term stability over a sequence of experiments in a stationary world". This material covers, amongst other topics, two trial product experiments, Bayesian inference and the general concept of conditional probability; independence of experiments being shown up as a special case. It is possible to model all of the pertinent relations on a two-fold product of STATLAB, which exists and is operational and which necessarily contains a non-trivial one trial storage capability, since the result at trial $\tau + 1$ may or may not depend upon a result at trial τ. Corresponding to the Cartesian product of experiments, the abstract metric side of the STATLAB facility is equipped with the multiplicative product (to realise the fundamental rules (for results α, β at trials τ, $\tau + 1$) namely $P(\alpha \text{ and } \beta) = P_\alpha \times P_\beta$ if and only if α and β are independent; otherwise $P(\alpha \textit{ and } \beta) = P_\alpha \times P_{\beta/\alpha}$.

The same topic relations ramify into other fields of knowledge; amongst these we have primarily dwelt upon simple probabilistic automata (i.e. finite automata with probabilistic state transitions) and the information theory in which these automata appear as stochastic, generally Markovian, sources. The supporting topic relations, entailed by "probabilistic automata" are to do with deterministic finite automata.

2.3. *Automaton facility*. A facility for modelling finite state machines, both in terms of standard components and of state transition graphs has been constructed and pilot studied. For machines of an interesting size the state representation is both tortuous and impracticably large (it is used only to establish the standard correspondences between finite automata and the grammars of various formal languages). The brunt of modelling is borne by the component part of the facility.

Using the well known correspondence between finite state machines and serial programs the component part of the facility is isomorphic to a programming language. Thus, instead of describing our own equipment we shall thus refer to Papert's programming language LOGO; which has been widely and successfully put to use. We concentrate upon one application of LOGO, computational geometry (LOGO may be used for writing programs, alias models, in almost any field of interest; for example, a semigroup can replace the infinite group of computational geometry).

2.4. *Computational geometry.* LOGO is a computer programming language clear enough to be interpreted and used even by children. Given any specific application of LOGO, boundary conditions are imposed upon the LOGO programs that may be written and executed; these are conveniently built into the appropriate peripheral equipment. For example, in its application to computational geometry, Papert (1970) makes extensive use of a computer controlled "turtle"; optionally a real machine (reminiscent of Gray Walter's robot like "tortoise") that moves on the floor, or an "imitation turtle" displayed on a screen. This creature is defined by a head end, a location in a plane, and a direction and distance of movement. Obeying instructions received when the program is executed, it describes a path in the plane and leaves a trace of this path. The boundary conditions are various inoffensive constraints upon the amount and direction of turtle motion, the fact that there is but one turtle in one state at once and the fact that this state is in register with the program state.

Any computational geometry program is a finite state machine. The trace is its state trajectory, over steps corresponding to the clock units, τ, of STATLAB. The program itself is a model; the program (which may not yet be satisfactory) that is written and chosen for execution at a given instant, t_0 corresponds, in our notation to the model $\{ x \}_{t_0}$.

2.5. *Comparative discussion.* LOGO and STATLAB are designed for different subject matters and frames of reference but have a great deal in common. The computational geometry student states a (tentative) goal for which he intends to write a program, namely, a relation in the domain of computational geometry. Suppose, for instance he wishes to draw "stellate figures with 6 points", this is a relation. He models the relation by writing a program which, on execution, causes the turtle to describe certain figures. If it does this job, the model is correct but not necessarily complete. It would be complete and correct if it does so for all parametric variations allowed by the restricted form of LOGO (for example, all permissible sizes of figure; all permissible regularities, and so on).

A student of computational geometry usually selects his own goals, but he does so under a global (more precisely, an L^1) description of what can be achieved in this field and often in the light of guidance from his peers or mentors. The implied permission is important; though, because of the project objectives, it is expressed in natural language. The implicit constraints would become explicit permissions in CASTE where they would be represented as a possibly open ended entailment structure. Both model making and LOGO programming operations can be represented as the construction of a series of finite state machines. Generally, there are indefinitely many ways of satisfying or bringing about a good (R_i) and the student is able to recognise, for himself, whether or not this model (program) is satisfactory by executing it and noting whether or not it works.

LOGO is more general than STATLAB (which is custom built and purpose oriented). But the main distinctions are due to the fact that STATLAB is incorporated in CASTE whereas LOGO is not (though it readily can be, given the requisite entailment structure; albeit, open to enlargement using the course assembly heuristic). In spirit, the systems are very similar.

2.6. *Blatantly heterogeneous modelling facilities.* Plate 11 is an entailment structure for a conversational domain of interest in medical education, "Diseases of the thyroid gland". If this subject matter is presented to a learner who is required to diagnose the maladies in question and non-verbal explanations are to be elicited or tutorial demonstrations employed, then it is necessary to have at least two modelling facilities and, for preference, rather more of them. These facilities are outlined in Plate 12. One modelling facility is a simulation, with free (student adjustable) parameters of the thyroid regulatory system. A model is an assignment of parameter values (for example, to "iodine intake" and variables that cut or establish chemically mediated couplings between the pituitary and the thyroid gland such as the variable "TSI level"). Execution of the model over trials τ, $\tau + 1$ gives rise to a state trajectory (a state, at τ, is the conjoint values, at τ, of the variables noted in Plate 11) and the student is encouraged, by demonstration, to build models that depict pathological conditions. The other modelling facility (really a cluster of facilities) is centered upon a decision theoretic structure due to Taylor and his colleagues: it is a means of choosing optimum clinical tests and/or treatments, on the evidence available about a patient suffering from a metabolic disorder involving the thyroid regulatory system.

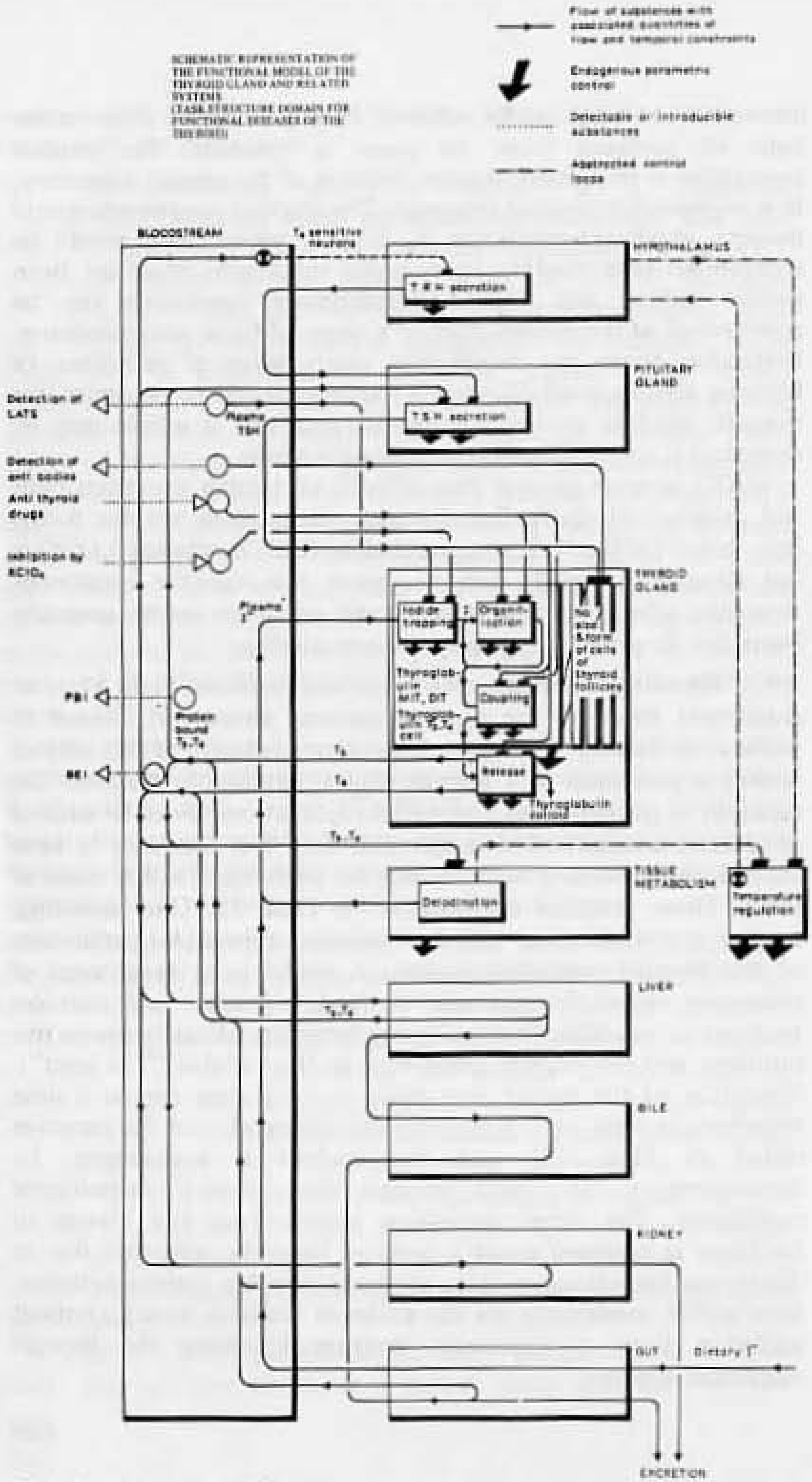


Plate 12, legend see p. 361.

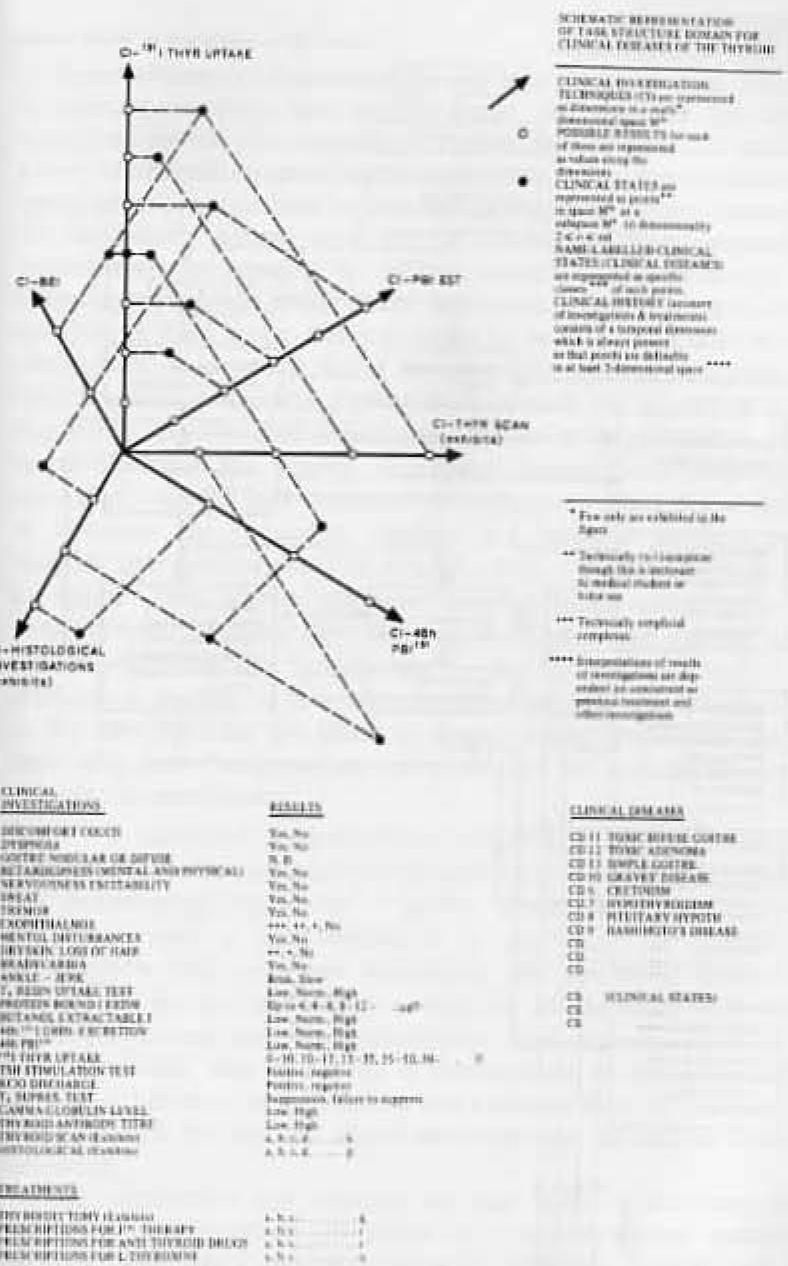
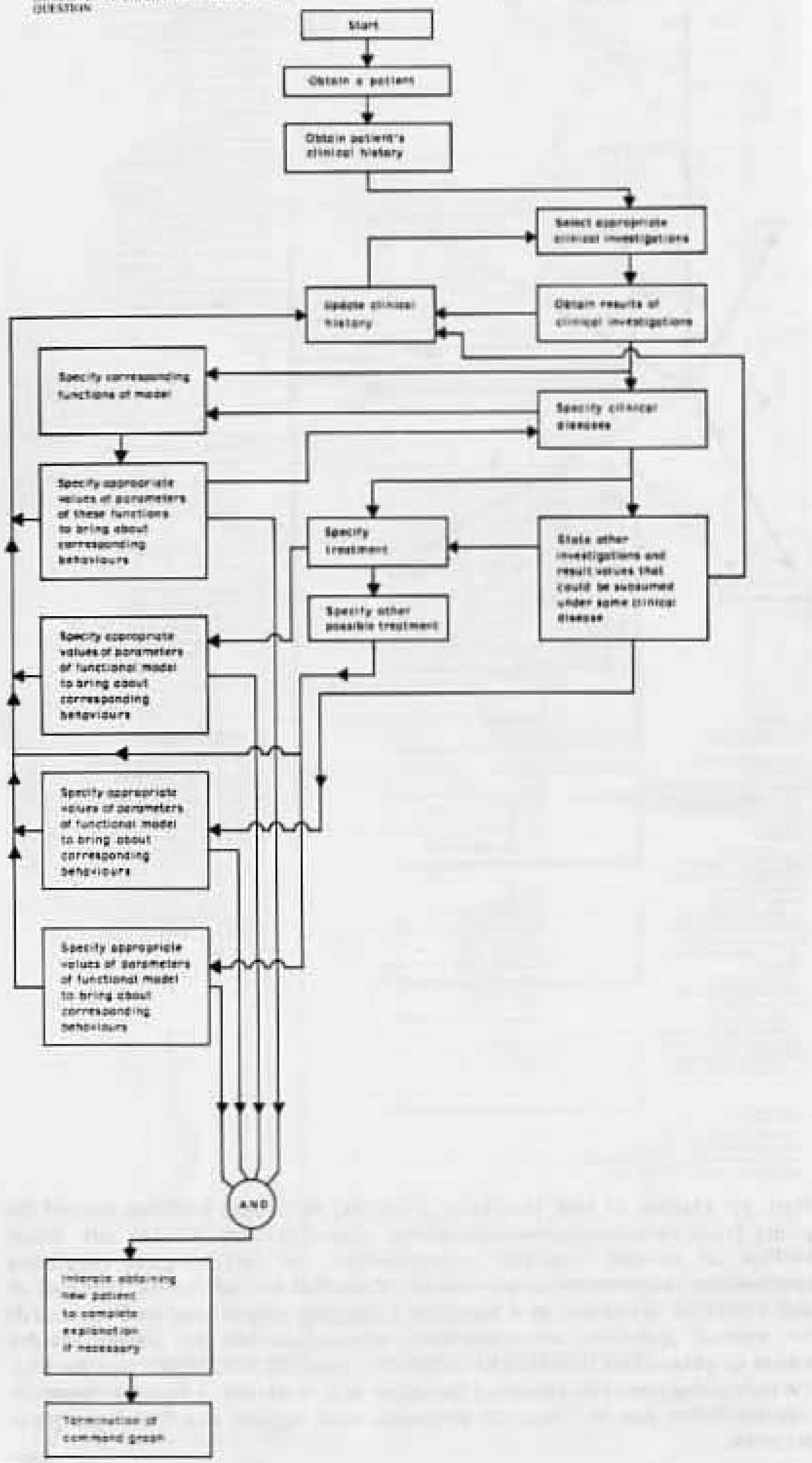


Plate 12. Outline of task structures governing modelling facilities needed for giving (non verbal) explanations of the topics in Plate 11. On left, block outline of control theoretic representation for physiological regulating mechanisms (parameters under control of student are indicated). This part of task structure is realised as a monitored analogue simulation; student models for normal operation are subsystems; student models for disease involve values of parameters that lead to instability, trapping conditions, and the like. On right, diagrammatic sketch of the other task structure; a decision theoretic representation due to Taylor (reference in text) realised as a digital computer program.

EXAMPLE OF TASK STRUCTURE AS REALISED DURING TUTORIAL CONVERSATION 12, AS DEMONSTRATION OR ANSWER TO EXPLANATION DEMANDING QUESTION



Quite obviously, the modelling universe of physiology/anatomy is distinct in kind and quality from the universe of clinical tests/treatments. The disparity is highlighted by noticing that only a few of the pathological state trajectories in the functional model (notably those that lead to limiting or trapping states) correspond to universally agreed and named diseases. Other pathological trajectories correspond to diffuse clusters of symptoms (the literature is replete with them) and their appearance, as sampled through clinical tests, depends upon how and when the tests are performed. Conversely there are some symptoms (for example, Myxodaema or exophthalmus) that cannot be modelled in the functional facility as it stands (any essays in this direction would involve adding the known functional connections between the regulatory system and abnormal subcutaneous tissue development, or between the regulatory system and possible mechanisms to mediate eye protrusion. The crucial point is that mere addition of a "signal" like "protruding eyes" is illegitimate according to the construction rules we have discussed). Thus, some models in the functional modelling facility bear a one to one or many to one relation to models in the clinical modelling facility; other models in the two facilities are many to many related and in parts of the field the two facilities are only related in a cognitive sense; namely, by entailment.

2.7. *A comment on behavioural objectives.* In this connection, it is profitable to compare the present approach with the doctrine of "Behavioural Objectives". Clearly, a modelling operation is a behaviour and a demonstration is one kind of behavioural prescription (the converse statements are obviously false; some behaviours are not modelling operations just as some behavioural prescriptions are not demonstrations). One extreme form of the doctrine holds that knowing is tantamount to behaving (latent behaving, perhaps) and a slightly less extreme form of the doctrine holds that all we can say about knowing may be said in terms of behaving.

Such arguments are difficult to deal with in the case when models/demonstrations are made in a homogeneous modelling facility, for the counter-argument involves much of the superstructure erected in earlier chapters. But the essential point is easily grasped in the context of several heterogeneous modelling facilities and a skill such as diagnosis. It is either necessary to maintain (as we did in the last section) that cognitive or entailment relations exist that are quite distinct from behaviour generating relations, or it is necessary to invoke a rather implausible principle of contiguity and to maintain that teaching diagnosis boils down to rehearsing certain disparate behaviours (using the functional and the clinical modelling facilities) in close temporal or spatial proximity.

Whereas task structures dictate behaviours (if given an imperative interpretation) the imperative interpretation of an entailment structure (or a path in it, even) only prescribes a "behaviour" in the strained and crassly extended sense of "the behaviour of an entity acting to modify its own (cognitive) structure".

3. Games and Control Situations

For subject matters like economics or industrial management, the laboratory is a microcosm in which decisions are taken and plans are made. A student may apply economic principles in regulating a simulated economy which is perturbed by random or partially lawful disturbances; alternatively, he may take part in a business game, pitting himself against a more or less rational opponent. Very similar comments apply to history as a subject matter except that, in this case, the student is presented with a simulation of conditions that held in the past and is asked to compare his course of action with the actual actions taken by the politicians or administrators of the period. Depending upon the context, this kind of exercise will more or less closely resemble a game or a control operation; as a limiting case, it becomes a modelling and problem solving activity of the type considered up to this point.

In the present section, we discuss the construction of game like and control like situations within the compass of a one clocked modelling facility (the constraints of Chapter 5, Fig. 1, as before). To represent these situations it will be necessary to introduce further icons or ordered sets of icons, though these superstructures collapse into Icon 10 when, in the limiting case, the game like or control like situation is reduced to a problem solving situation.

Some caution is needed, because the participants, A, B are not necessarily the players; they may be, of course, and, if so, this fact will be stated. Clearly (since they are also in conversation) A and B cannot always be game players, for the act of playing depends

upon isolation apart from the moves made. Moreover, even on those occasions when A and B do engage in play, it is possible that they nominate or construct protogonists who actually do the playing for them. (So, for example, if B is a teacher, he may select an appropriate simulation game for student A to play, though he does not play against A himself.)

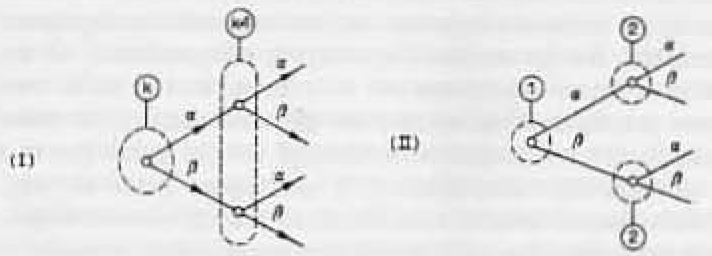


Fig. 9.1. Games in extensive form. Nodes represent choice points; the directed arcs are labelled by the names of moves $\alpha, \beta, \ldots$ (different moves if the arcs emerge from one choice point). Each choice point is assigned to 1 of $1, \ldots, k, \ldots, m$ players in prescribed playing order. Partitions of choice points of a player into disjoint sets are known as "information sets", designated by dotted lines, and interpreted thus: "a player cannot distinguish between choice points in one information set". If two or more choice points belong to one information set (in dotted lines) the emergent arcs bear the same names. No path enters an information set more than once.

3.1. *Standard games.* The development of standard two person game theory goes as follows. The two players, who are isolated apart from the interactions due to playing, are each able to select from a finite number of moves at each step in a game. The possible outcomes of play are depicted as a tree (Fig. 1). It is assumed that each player knows this tree, unless the rules dictate concealment of some prior moves; in which case, a choice of move is made in the absence of complete information about the player's current locus in the tree. The terminal branches of the tree (the outcomes) are associated with a payoff, θ_a, θ_b to each player a, b. If it happens that for each outcome the gain to one player equals the loss to the other player then the game is constant sum. The payoff may either be determined by ranking the players preferences over the outcomes or by edict (each player is told that so much of a commonly worthful commodity is assigned to each outcome). In either case the standard game theories prescribe rational action i.e. that each player aims to maximise his expected payoff. For example, if the game is constant sum, then a completely avaricious maximisation is rational.

Any particular play of the game is represented as a path through the game tree; conversely, (since the players are isolated apart from the play) they could each have stipulated, beforehand, a player strategy; a set of statements asserting the move that will be selected at each choice, contingent (perhaps) upon prior moves and, thus, upon a players location, at this moment, on the game tree. Thereafter, the game could be played independently of the players. For a finite total number of moves, and a finite tree depth, there is a finite number of strategies and play of the game can be substituted by choice of a strategy (or for some games a weighted probabilistic combination of strategies) prior to play, together with the automatic execution of the chosen strategies. The normal representation of a two player game is thus a matrix in which the rows correspond to one players strategies, the columns to the other players strategies and the entries in which are values of the payoff function.

3.2. *Games as automata.* In any actual game there are recognisable situations. For example, in a board game, the situations are configurations of playing pieces. According to the game tree formulation, any recurring situation will be counted as distinct, and although the move sets may differ from branch point to branch point they are preordained. As a result of this, a strategy cannot involve conditional statements or, to rephrase the matter, it is a feedforward operation; feedback is not comprehended. An alternative formulation, due to Banerji (1970) avoids this restriction upon the use of feedback and stresses the fact that though a set of move-like "operators" is given, an "operator" may or may not be applicable to the prevailing situation.

A set, C, of situations $c \in C$ is taken as fundamental. Certain of the situations are distinguished as outcomes $c \in C^{\boxtimes} \subset C$ and the outcome set is partitioned into disjoint subsets; for example the sets C^{+}, C^{-}, C^{X} where, if $c \in C^{+}$, then player a wins and player b looses if $c \in C^{-}$ then player b wins and player a looses and if $c \in C^{X}$ the game is a draw, i.e. C is the set of those states to which no operator a_i or b_j (defined below) is applicable and which are not labelled + or −. The players are equipped with sets of moves applicable in certain situations only which (as in Chapter 5, Section 1) we call operators including the null operator (a_i for player a and b_j for player b). Operators are described, in the abstract, as functions from a subset of situations onto a situation. Let λ_a be the set of a_i and λ_b be the set of b_j and let C_a be the set of situations to which $a \in \lambda_a$ (player a moves) are applicable and C_b the set of situations to which $b \in \lambda_b$ (player b moves) are applicable. A play of the game is any path starting at some initial situation, $c \in (C - C^{\boxtimes})$, in which operators a_i and b_i are applied, either in turn or in a synchronised manner, so that eventually $c \in C^{\boxtimes}$ (when, by previous definition, either $c \in C^{+}$ or $c \in C^{-}$ or $c \in C^{X}$). At this point, the game ends, the outcome being announced to player a, and player b. Since player a and player b, are isolated apart from their play and since their moves are synchronised, any play they might execute could be expressed, beforehand by a strategy which, if precautions are taken to ensure that each game has a beginning and an end, can be represented as a function Strat from $C - C^{\boxtimes}$ into λ_a or λ_b.

That is:

Strat a: $(C - C^{\boxtimes}) \rightarrow \lambda_a;\ a_i \in (C_a \times C)$ for $a_i \in \lambda_a$:

at least one $a_i \in (C_a \wedge (C - C^{\boxtimes})) \times C^{\boxtimes}$

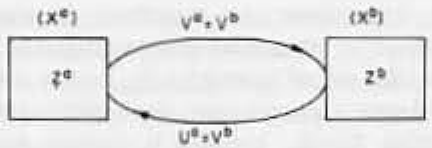
Strat b: $(C - C^{\boxtimes}) \rightarrow \lambda_b;\ b_j \in (C_b \times C)$ for $b_j \in \lambda_b$:

at least one $b_j \in (C_b \wedge (C - C^{\boxtimes})) \times C^{\boxtimes}$

where any a_i, b_j, in λ_a, λ_b, satisfy a binary "legal move" relation Γ. From this, it will be evident that the entire game is representable as a finite automaton and that Strat a and Strat b are synchronously coupled finite automata (signified "o") forming the automaton. (Strat a) o (Strat b). Hence, Strat a and Strat b may be realised as models $\{x\}^a$, $\{x\}^b$, in the modelling facility already specified and, if so, their one execution under the τ clock is a play

$$c_\tau = a_\tau(b_\tau(a_{\tau-1}(b_{\tau-1}(\ldots(a_1,(b_1(c_0)))\ldots))))$$

if the τ clock is zeroed at the end of each play ($\tau > 0$ throughout play). c_τ, in this expression, is the instantaneous state of play, which contains the data available to Strat a and Strat b at trial τ. This data can be augmented by a record of states up to τ namely $< c_1 \ldots, c_\tau >$.



(V) = Output states

(U) = Input states

(Z) = Internal states

Fig. 9.2. Representation of a game in terms of simple automata.

One construction for the finite state machines (alias the strategies) in terms of the existing notation (u, input states; v, output states; z, internal states) is shown in Fig. 2; the instantaneous states correspond to internal states z of $\{x\}^a$ and $\{x\}^b$. The state graph of the joint automaton $\{x\}^a \circ \{x\}^b$ is the directed graph of the game, in which nodes represent situations and directed arcs represent the application of operations.

Banerji (1970) is primarily concerned with effective models of game playing; for example, devising winning Strat a that (for some τ) place $c \in C^+$ regardless of the moves due to Strat b. The interested reader is referred to the original. One special point, that bears directly on the immediate discussion, is that no essays in the direction of optimality are likely to yield a useful result unless pertinent subsets of the set C are described concisely. For example, Banerji considers the existence of nested "evaluating subsets" $E_r \subset C$ (Fig. 3) such that $E_0 = C^+$ and for each r = 0, 1, ... there is some (applicable) a that for all applicable b yields $a(c) \in E_{r-1}$ if $c \in E_r$ (that is, the E_r lie at an increasing "distance" from the desired outcome and an automaton can be given the "goal" of minimising this "distance"). Call a distance minimising operator a goal directed operator, and, as a further example, suppose a situation, c, in which there is no goal directed operator; then it may be possible to set up a subgoal (under the main goal) to transform C into C*, such that C* is in some E_r, so that a goal directed operator is applicable to C*.

In terms of our own formulation, it is possible to write a reasonable programme for $\{x\}^a$, if and only if, there is an

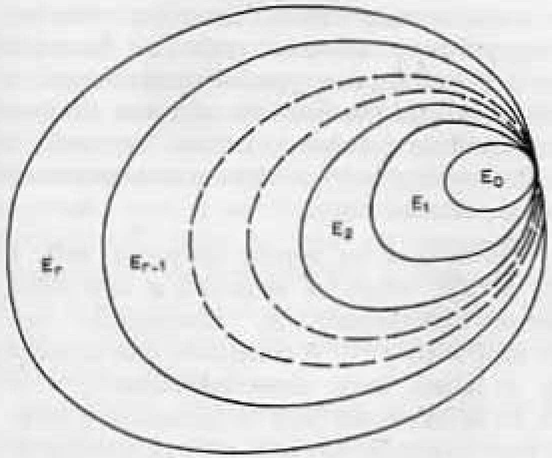


Fig. 9.3. Nested "evaluating subsets".

effective description. But in this formulation C is constructively specified in terms of those relations regarded, at an occasion n, as properties and consequently as predicates in L^0 (i.e. describability, in Banerji's sense, is tantamount to a well partitioned state set X_A). Within this formulation, player b is known to player a in terms of what players can do (the relation Γ on C) and in terms of the outcome sets C^+, C^-, C^X.

3.3. *Game strategies, control strategies, and problem solving.* As noted earlier, very similar comments apply to any control situation. Strat a is a control intended to counter the action of any disturbance or perturbation strategy namely, Strat b. The isomorphism between game playing and control was first pointed out by Ashby (1964a), in the context of his law of requisite variety, which finds statistical expression in Conant's (1968) theorem (that the degree of regulation available to Strat a depends upon the variety of a in λ_a and upon the selective information content of the strategy class). The isomorphism exists just because feedback as well as feed forward is countenanced in game strategies, as it is in most control strategies. Banerji (1970), thinking along very different lines, has axiomated both board games and control situations as special cases of a "Marino situation" (using the technical apparatus sketched in the last section). Further, if the set C^X is empty, if Strat b has but one move to establish an initial configuration, then the "Marino situation" reduces to problem solving, (this is Banerji's "Windeknecht situation") of exactly the type we have referred to as modelling or explanation within the confines of a one clocked facility. "Play" is the execution (under τ) of the "model"; a sequence $< v_1 \ldots . v_j >$. Possibly $v \Leftrightarrow z$ when the sequence is $< z_1 \ldots . z_j >$ (Chapter 5, Section 1).

3.4. *Development of Icons.* Let us equate player a with a procedure class, $\underline{\mathrm{Proc}}^0_A$ i in π^0_A which, at occasion n in a strict conversation, characterises participant A. Distinguish two possibilities; namely (1) that with $\tau = 0$, A constructs one or more models that are Strat a (alias, finite state machines $\{ x \}^a_t$ submitted, at instants t, to be executed with t constant over trials $\tau = 1, 2, \ldots.$) and (2) that A chooses amongst a set of alternative strategies available to him at some instant $t = t_o$, the chosen Strat a being executed over trials $\tau = 1, 2, \ldots.$ with t constant; only one choice is permitted, but it may express A's degree of belief by a confidence estimate taken at $t = t_o$ about the best Strat a; this probabilistic statement being used as a bias to an otherwise random strategy selecting device.

Of these possibilities, (1) furnishes A's non-verbal explanation of how he intends to bring about a relation R^a_i in which the index i characterises the game situation at the occasion in question and the superscript indicates the fact (of rationality) that A intends to place $c_T \epsilon C^+$ for some J regardless of Strat b. In contrast (2) is the selective answer (albeit a probabilistic answer) to a multiple choice question with an alternative set consisting in the possible strategies.

If the situation is symmetrical, player b is equated on occasion n with $\underline{\mathrm{Proc}}^0_B$ i in π^0_B (participant B's L^0 repertoire); Strat b is either constructed (case (1)) or selected (case (2)) and the aim is to execute Strat b to bring about a relation R^b_i in which $c_T \epsilon C^-$.

Given case (1) it is always possible to obtain case (2) by the expedient of setting the t clock at $t = 1$ and presenting a finite set of alternatives (some or all of which would have been generated as Strat a). This gambit is employed to construct multiple choice situations in Chapters 5 and 6. But the difference between case (1) and case (2) also highlights a further (and not an immediately obvious) distinction between games regarded as automata and standard games.

A standard game theory is a normative theory; not a descriptive theory. It makes no comment upon the psychological mechanisms of game playing. It is, in fact, primarily concerned with the idea of a game solution (for example, the well known "Minimax" solution, of Von Neumann and Morgenstern) or, of greater immediate convenience, a Nash, or equilibrial solution. That is, suppose a real or an imaginary iterated play of a game (so that a game is not really a "one off" business) and a norm such as "rationality" or "competition in a game of pure skill"; Strat a and Strat b are the (not necessarily unique) "equilibrial solution" if neither player gains an advantage (under the norm) by deviating from these strategies.

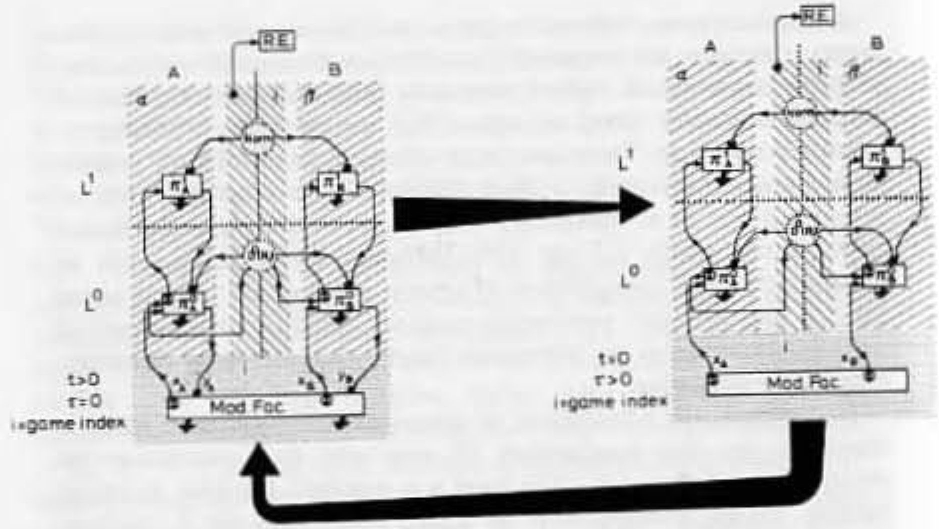
In contrast, the conception of games as automata does make a comment on the mechanism of play and this mechanism is exhibited insofar as $\{ x \}^a \equiv$ Strat a is a model for how A plays; further, model construction by $\underline{\mathrm{Proc}}^0_A$ i (see Chapter 5, Section 1.7.5.1 for its symbolic expression) is, in Banerji's nomenclature, a search for a strategy Strat a.

Even though standard game theory makes no comment upon mechanism, it is, all the same, possible to impose the standard game as a normative constraint (like an experimental contract) upon an L conversation and to use sequences (playing cycles) of icons to represent the conversational events that would, of necessity, accompany any play of the game. Conversely, games in the sense of automata are part of the conversational process insofar as it entails the construction of models $\{ x \}^a \equiv$ Strat a and $\{ x \}^b \equiv$ Strat b.

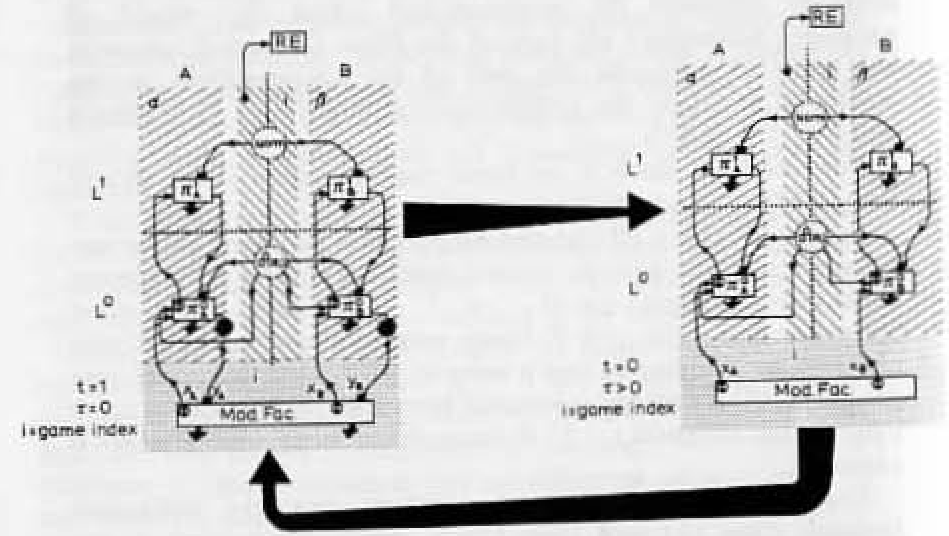
For example, there is an iconic representation (Icon 19 and Icon 20) for any constant sum game.

In case (1) (Icon 19) the playing cycle has two parts during any occasion, n. First, with the t clock operating and $\tau = 0$ the players construct automata $\{ x \}^a$, $\{ x \}^b$. Next with the t clock constant, but with $\tau > 0$ these automata are executed and observed by the players over a series of trials $\tau = 1, 2, \ldots.$ In case (2) (Icon 20) a strategy is selected from a set of alternatives at $t = 0$ (as a result of which $t \rightarrow 1$). Subsequently, the selected strategy is executed.

From a syntactic or formal point of view, the distinction between Case (1) and Case (2) is fairly unimportant if the situation is genuinely symmetrical. In practice however, symmetry is unusual. Though A commonly acts over occasion n in the



Icon 19.



Icon 20.

restricted role of player a, he is, by virtue of his experience in this role, learning the role or "learning to play the game". Under these circumstances, B is a teacher or a tutorial heuristic. B selects <u>Strat</u> b (in itself the strategy of a purely competitive player or, control wise, a pure disturbance) so that A has a reasonable chance of winning/maintaining stability. The earlier examples of adaptive regulation using the steady state technique are typical of B action.

3.5. *Icons for partially co-operative interaction.* Standard game theory (somewhat uneasily) comprehends various "partially co-operative" game situations, and there is a body of literature concerned with the nature of a "solution" to such a game; for example, Luce and Raiffa (1959). Partially co-operative games are realised insofar as the (normal form) payoff matrix contains entries such that Player a and Player b can gain more by joint or concerted action (one form or another of mutualism) than they can independently. That is, for iterated play, the average value of payoff obtainable by appropriately correlated moves exceeds the maximum average value obtainable by moving independently.

If cooperative interaction is countenanced, there is a world of difference between Case (1) and Case (2) of the (last) Section 3.4. Actual iteration of play allows for covert communication; Icons can be drawn to represent this fact; Icons may also represent overt communication in language L^0. We do not present these.

Whilst recognising the syntactic utility of game theory as it stands, we are uneasy about the game theoretic representation of a norm, of knowledge of another player, and of the pervasive idea that the players actually know the preference orderings or values over all outcomes. In fact, we doubt whether players have "preference values" except for the "preferences" latent in the persons they are. Hence, in the next Section, we take a different tack and represent the system in terms that make no special appeal to assumptions about preferences, prior agreements, rational norms, payoffs, utilities and so on.

In a certain sense, between-player interaction is strongly presupposed; the interaction might, under special circumstances, involve actual bargaining (though if the game were iterated, it would be represented, in tabular form, as a different game). In general, the pertinent between-player interactions of this (non standard) approach are conceptual; one player entertains correct anticipations or hypotheses about the other and vice versa, perhaps. The (non standard) theory deals with cases where the anticipations, hypotheses, or actual acts of communication are correct. But the psychological theory obtained by adjoining it to conversation theory accommodates other possibilities.

4. Identification with non-standard game theories

A game theory is "non-standard" if the players are not required to subscribe to a fixed norm (for example "rational and avaricious") but may, in a manner that is still represented within the theory[23] and allowed by the normative contract of the game, construct and modify their hypothesis about one another; i.e. if they can contemplate alternative hypotheses within the theory or the contract on which it is based. For example, Braithwaite's (1955) arbitration games may be regarded as non-standard (Pask 1963) though this interpretation is entirely optional. The present discussion is confined to the class of metagames (Howard 1966, 1973). Metagames are non standard in the sense already outlined and are general enough to comprehend all standard games as special cases. Moreover, they are founded upon the simple and lucid idea of rationality which underpins the original work on game theory (Von Neumann and Morgenstern 1953) stripped of the additional and often less plausible assumptions with which the theory was ladened during its later development. Given a set of alternatives from which to choose, a player is rational if he selects the preferred alternative. As Howard (1973) points out, human beings are not generally rational. Under some circumstances several people cannot be jointly rational; under others, two people fare better if they are not rational; finally, there are cases in which an "obviously" rational choice is stupid.

4.1. Historically, the theory of metagames (Howard, 1973) arose in response to the second circumstance (two people fare better if they are not rational). The matter is discussed in an article by Rapaport (1968) noting that this situation is ubiquitous

[23] "Within the theory" is important. Obviously real participants can always do so if they wish. The point is that the normative contract established by a non-standard theory is liberal enough to admit their doing so without breaking the rules.

Fig. 9.4. The outcome matrix for "Prisoner's Dilemma" construed in normal form. Only rank ordering of "payoff values" are of significance. The entries $<X, Y>$ indicate payoff to $<$Player 1, Player 2$>$ as a result of the specified outcome.

and counts as a general paradox of gaming. The situation is exhibited by the game of "Prisoner's Dilemma" which, in normal or tabular form, is characterised by the outcome matrix of Fig. 4 (the payoff values signify ranked preferences only). The solution, which is also rational in the standard game theoretic formulation, is the outcome with payoff $<2,2>$. But this outcome is counterintuitive, since if the players entertained differential hypotheses about each other or agreed to cooperate they could gain more, the outcome with the $<3,3>$ payoff.

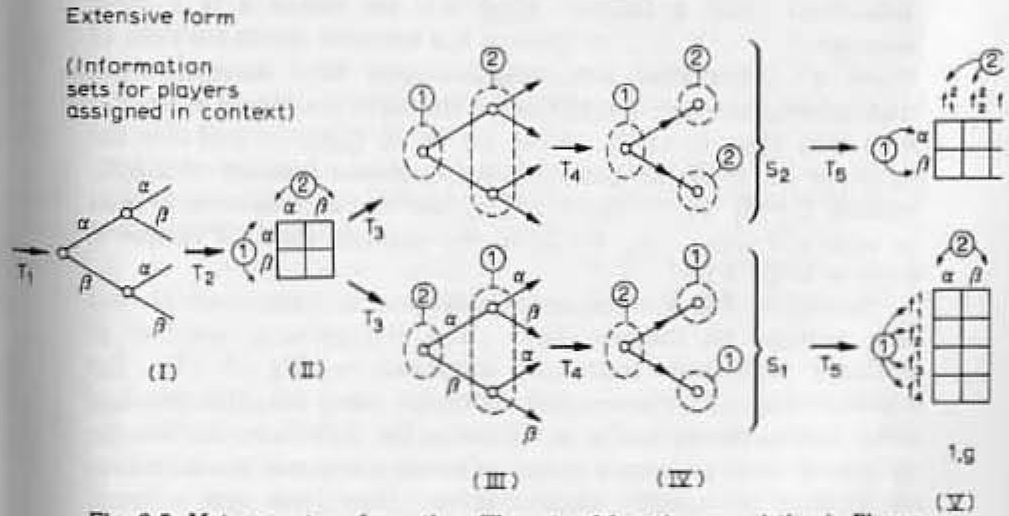


Fig. 9.5. Metagame transformation. The sets of functions consisting in Player 2's choices in 2, g and Player 1's choices in 1, g are as folows: $f_1^2 = \alpha$ if α; α if β; $f_2^2 = \beta$ if α, β if β; $f_3^2 = \alpha$ if β; β if α; $f_4^2 = \beta$ if α, α if β and $f_1^1 = \alpha$ if α, α if β; $f_2^1 = \beta$ if α, β if β; $f_3^1 = \alpha$ if α; β if α; $f_4^1 = \beta$ if α; α if β. These are the conditional strategies of 1st order metagame.

4.2. To show the structure of a metagame in which this "dilemma" can be resolved consider the game tree in Fig. 1 (II) and the series of transformations (T) in Fig. 5. For notational convenience the players are relabelled "1" and "2" (there being m players in general) rather than "a" and "b", as proposed in Fig. 1 (II). Player 1 has moves α and β: Player 2 has moves α and β. It is certainly true that there are 4 results of play: equally, if Player 2 moves after Player 1 (as proposed) and in ignorance of Player 1 (that is, devoid of consideration of how Player 1 may have chosen) there are 4 and only 4 move combinations. But suppose that Player 2, who moves after Player 1, correctly guesses, actually knows, infers or anticipates how Player 1 will move and moves according to this correct guess, knowledge, inference or anticipation, then Player 2 may select from 4 contingent plans namely α if Player 1 plays α: α if Player 1 plays β; β if Player 1 plays α and β if Player 1 plays β. Conditional plans of movement are just as much strategies *if* Player 2 *does* rightly consider the actions of Player 1, as the uncontingent plans set out in ignorance of Player 1's intention. Thus Fig. 1 (I) differs from Fig. 1 (II) only in the placement of the information sets of Player 1 and Player 2 and it is convenient, on this account, to adopt the following definitions: that a strategy Strat k i, for Player k in a game amongst 1 . . . . k . . . . m players is a function which for each of Player k's information sets, assigns a move label emerging from that information set; that the set of strategies for Player k, say Z_k, such that Strat k, i ϵ Z_k, is the set of all Strat k,i and that the outcome set Z of the game g is the Cartesian Product of the Z_k namely $Z = Z_1 \times \ldots . Z_k \times \ldots . Z_m$ so that an outcome, z, is an m-tuple $z = \langle z_1, .. z_m \rangle \epsilon Z$. In the example cited, of course, Z reduces to $Z_1 \times Z_2$.

Turning to Fig. 5, a 1st order metagame is constructed around this example as follows. First, (T_1), the game is specified in ordinary extensive form, as suggested in Fig. 5 (I). The transformation T_2 carries this extensive form into the standard form that is shown in Fig. 5 (II) using the definition of outcome to specify cells or entries in the outcome array and the definition of strategy to specify its coordinates (here, row and column, only). From this normal form we obtain, by T_2, a special extensive form of the game (Fig. 5, III) characterised by the players moving in a given order in ignorance of the moves of previous players; in general, there are m! such orders; as shown 1,

2 and 2, 1 only. The "in ignorance" clause is satisfied by assigning appropriate information sets. From any one special extensive form we obtain one special normal form which is a particular 1st order metagame (Fig. 5, IV) here the metagames 1, g and 2, g by using transformation T_4 which replaces the information set surrounding the last player's choice points by as many separate information sets as there are choice points (to give, on interpretation, a situation in which the last player correctly knows or correctly infers the results of all the moves of the preceding players) and T_5 which labels this set of situations as S and replaces the alternative set of moves available to the last player by a set of functions F_k. These functions (the collection $f_{ki} \epsilon F_k$) become the last players conditional moves or metagame strategies and the Cartesian product of these with the other players strategies constitute the 1st order metagame outcomes (the metagames 1, g and 2, g are exhibited in Fig. 5, V).

Now the last two transformations may be reapplied recursively. For example, if, in 2, g, Player 1 conceives Player 2 as entertaining

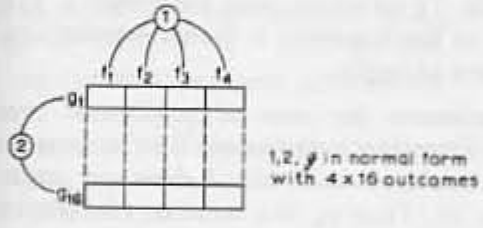


Fig. 9.6. The 2nd order metagame 1, 2, g. A tabulation of conditional strategies open to player 1 (they are the g_1 . . . . g_{16} functions in the set G) and the metaequilibria.

g_1 = α if f_1 ; α if f_2 ; α if f_3 ; α if f_4	g_9 = β if f_1 ; α if f_2 ; α if f_3 ; α if f_4
g_2 = β if f_1 ; β if f_2 ; β if f_3 ; β if f_4	g_{10} = α if f_1 ; β if f_2 ; β if f_3 ; β if f_4
g_3 = β if f_1 ; β if f_2 ; β if f_3 ; α if f_4	g_{11} = α if f_1 ; β if f_2 ; β if f_3 ; α if f_4
g_4 = β if f_1 ; β if f_2 ; α if f_3 ; β if f_4	g_{12} = α if f_1 ; β if f_2 ; α if f_3 ; β if f_4
g_5 = β if f_1 ; β if f_2 ; α if f_3 ; α if f_4	g_{13} = α if f_1 ; β if f_2 ; α if f_3 ; α if f_4
g_6 = β if f_1 ; α if f_2 ; β if f_3 ; α if f_4	g_{14} = α if f_1 ; α if f_2 ; β if f_3 ; β if f_4
g_7 = β if f_1 ; α if f_2 ; β if f_3 ; α if f_4	g_{15} = α if f_1 ; α if f_2 ; β if f_3 ; α if f_4
g_8 = β if f_1 ; α if f_2 ; α if f_3 ; β if f_4	g_{16} = α if f_1 ; α if f_2 ; α if f_3 ; β if f_4

The rational outcome $\langle 2,2 \rangle$ in Fig. 4 is obtained by the combination of $\langle f_2, g_2 \rangle$; the metaequilibria that are metarational for Player 1 and Player 2 and stable in 1, 2, g are obtained by the combination of $\langle f_3, g_4 \rangle$ and $\langle f_3, g_{12} \rangle$ and give rise to the outcome $\langle 3,3 \rangle$ in Fig. 4.

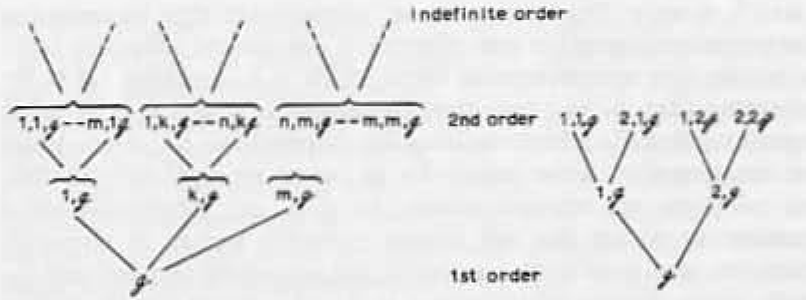


Fig. 9.7. The general hierarchy of metagames for m players and a construction for the special case of Player 1 and Player 2.

strategies $F_2 = \{f_1, f_2, f_3, f_4\}$ then Player 1 can construct doubly conditional strategies $G_1 = \{g_1 \ldots . g_{16}\}$ that are functions from F_k (here F_2) to his moves. These are tabulated for a metagame 1, 2, g, in Fig. 6. The Cartesian Product of $F_2 \times G_1$ determines a 2nd order metagame; namely, 1, 2, g. In general, there is an infinite hierarchy of metagames (Fig. 7), of which only the forms 1, 1, g (or k, k, g) are trivial. Part of the hierarchy is shown exhaustively for the case of the two players example.

4.3. Howard (1973) introduces the idea of preference over outcomes as a function ψ_k from the outcome set Z of an original game g (specified for the kth player) such that k does not prefer $\psi_k(z^0)$ to z^0 (for all $z^0 \in Z$). Thus ψ_k is a function (for player k) from Z to the power set of Z (the set of all subsets of Z). It may, for example, be represented as a graph in which nodes depict the possible outcomes and the directed arcs point, for each outcome, to those that are not preferred.

As special cases ψ_k may be an ordinal function (player k's rank ordered preferences; for example, those shown in Fig. 4 for the θ payoffs of Section 3.2), but it is important in the sequel to preserve the generality of the original formulation and it is legitimate to conceive ψ_k as a graph.

A game G is thus a 2m-tuple consisting in an ordered set of strategy sets (for m players) and an ordered set of preference functions (for m players), namely

$$G = \langle \langle Z_1 \ldots . Z_k \ldots . Z_m \rangle , \langle \psi_1 \ldots . \psi_k \ldots . \psi_m \rangle \rangle.$$

It is crucial in the formulation of metagames that the players are not (as in a standard theory) supposed to know directly each other's "preferences". As Howard points out, this is, in fact, true of all real games (in contrast to quasi-games in which the players are told of numerical ranks, utilities, or whatever). More fundamentally a "preference" is represented as it empirically is; namely as a valuation propensity of a player; overtly stated or not. It characterises the player insofar as a particular domain of activity is concerned. We later equate it with one description of the concept a player has at the moment of play.

In any case, we concur with Howard entirely and reject the standard "decision theoretic" development, for empirical and operational purposes (that is why there has so far been no explicit mention of participant's preferences; they are, as will be argued, implied by our specification of participants).

An outcome is rational for player k (in the original and uncontentious sense) if it belongs to a set μ_k which is defined (on recalling that $z = \langle z_1 \ldots . z_k \ldots . z_m \rangle \in Z$) by

$$\mu_k = \{ z^0 \text{ in Z such that, for all } z_\nu ; \quad \psi_k \langle z_1 \ldots . z_k \ldots . z_m \rangle \subseteq \psi_k (z_0^0) \}$$

or, for the two player example, by

$$\mu_1 = \{ z^0 \text{ in Z such that, for all } z_1 ; \psi_1 \langle z_1 , z_2 \rangle \subseteq \psi_1 (z^0) \}$$

and

$$\mu_2 = \{ z^0 \text{ in Z such that, for all } z_2 ; \psi_2 \langle z_1 , z_2 \rangle \subseteq \psi_2 (z^0) \}$$

According to most forms of standard game theory an outcome z^* of play is stable if it is rational for all players; that is, $z^* \in \bigcap_k \mu_k$ (the intersection over k of the μ_k). The various predictions of metagame theory have in common the idea that stable outcomes of metagames are metarational; though neither necessarily nor usually rational (in g) for all players.

4.4. Metarationality is defined in varying contexts with more or less strength; only some definitions are exhibited.

Turn to Fig. 5 and notice that whatever (conditional) strategies are adopted (in the example by Player 1 and Player 2) the result of playing them will yield some outcome in the original game g. If

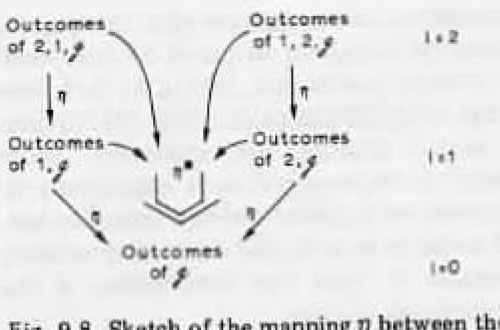


Fig. 9.8. Sketch of the mapping η between the levels in a metagame hierarchy and * the iterated application of η.

the players preferences are expressed over the actual results of play and are not influenced by the conditionality of the strategy employed to achieve them (that is, if ψ_k is invariant for player k, with respect to the order of the metagame) then it is possible to specify a function η^*, that carries the set of outcomes of any metagame derived from g onto outcomes in Z, the outcome set of the original game g. This is done by showing (Howard 1973) the existence of appropriate precedence and inclusion properties amongst the outcomes of metagames; by defining a function η that maps a metagame of order $\ell + 1$ upon a metagame of order ℓ and specifying η^* as the iterated application of η to yield outcomes in the original game g which is the metagame of order 0; the construction being sketched for a 2 player example like "Prisoner's Dilemma" in Fig. 8. In this case, a metarational outcome for any player k in a metagame (of an arbitrary order) is a member of the set.

$$\eta^* (\mu_k \text{ (metagame)})$$

or in particular, for Player 1 and Player 2, of

$$\eta^* (\mu_1) \text{ or } \eta^* (\mu_2)$$

yielding an outcome in g which is rational. An outcome is called general metarational for Player k if it belongs to the further set

$$\gamma = \bigcup_{\text{All metagames}} \eta^* (\mu_k) \text{ (metagame)})$$

Finally, an outcome is a general metaequilibrium if it belongs to the set $\bigcap_k \gamma$; it is general metarational for all players.

The matter can be phrased somewhat differently. In developing the theory of metagames Howard (1966) supposes that Player 1 is able to entertain and to choose between several hypotheses[24] regarding Player 1 and that Player 2 is able to entertain and to choose between hypotheses about Player 1. In other words the players have metahypotheses regarding how rationality is achieved and the metahypotheses are modelled as conditional strategies. But, regardless of how rationality is achieved preference invariance (the μ function and implicitly the η function) depends upon a fixed valuation of the basic results of play.

4.5. Although the theories are at first sight entirely distinct (and though they were, in fact, differently motivated, empirically tested and conceived) there is a great deal of similarity between Howard's metagame theory and our own theory of conversations; to the extent that results from metagame theory may be used to buttress conversation theory and possibly vice versa. Some of these similarities will be exhibited; others will be developed in the next volume (especially those to do with many person situations where the fields of application are close to each other). One of the few results we do employ (for the burden of the argument is structural, not mathematical) is Howard's theorem 1 (characterisation theorem) and its corollary 1. 2, to the effect that, for m players all metarational solutions are revealed by an examination of m! levels in the metagame hierarchy.

Before attempting an identification (which will, in any case, be no more than approximate) it is useful to clear up a point about realisability which all game theorists are bound to tackle somehow (if they deal with reality as well as mathematics) but which Howard and Banerji (Section 3.2) handle in rather different ways. Howard comments that the strategy set $Z = Z_1 \times \ldots . \times Z_m$ though it exists, is usually enormous; the game frankly cannot be represented in extensive form; it may even be infinite. Obviously,

24 Howard calls the hypotheses theories and the metahypotheses metatheories; the wording is changed to avoid confusion in the present context. The reader should also notice the aberrant usage of "meta"; thus a metahypothesis is not an external observer's hypothesis stated in L^* but A's hypothesis (about a class of L^0 hypotheses) stated in L^1. In contrast, the entire non-standard theory describing A and B is an external observer's metatheory, stated in L^*.

the metagame hierarchy is much larger than the game, so that (though it exists) it certainly cannot be represented in extensive form. Hence, in practice, Howard proceeds directly to the normal form using various expedients to elicit reasonable alternative sets. Banerji tackles the same problem (albeit in the case of structured games) by pointing out (Section 3.2) that a tractable set of game situations, C, (and, as a corollary, a tractable set of strategies) must be "easily described"; for example, one descriptive technique uses "evaluation sets" as shown in Fig. 3. In our own theory of conversations, a hybrid approach is employed when this problem becomes obtrusive (as it does in the next few paragraphs).

Conditions in a modelling facility or the modelling operations carried out must be L describable over the domain of relations (R_i) to be brought about or modelled; this, essentially, follows Banerji's approach. But the number of executions or explanations that might be used to represent R_i may still be indefinitely large and in equating these with the constructs of metagame theory, it is necessary to consider all of the ways in which executions or explanations may be truncated (by stopping the t clock, as in Chapter 5 and Chapter 6 to generate <u>Alt Sets</u> or to pose <u>PQuests</u>). The number of possible truncations is thus very large, and in winnowing them down to obtain a manageable sample we pursue Howard's approach, or something near to it.

With these comments in mind, notice that the outcome set Z (of a given game, *g*) may be identified with the outcome set C^{∞} of section 3.2. This seemingly outrageous statement is actually defensible; the incongruity between our interpretation of Banerji's formulation and Howard's formulation is due to a difference of perspective only.

From Banerji's point of view, a game is a mathematical (or a computational) entity for which an outcome set is ordained as the union of disjoint subsets C^+, C^-, and C^X. Our interpretation of Banerji's theory does no violence to the computational idea that a strategy (<u>Strat</u> a, <u>Strat</u> b) is a serially executed model, a finite state machine.

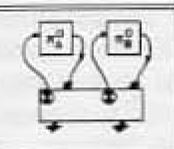
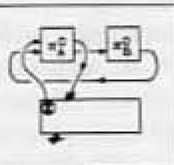
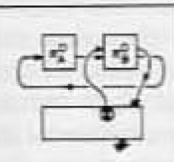
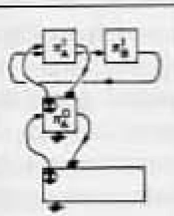
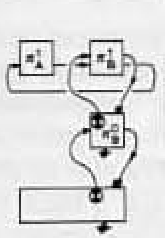
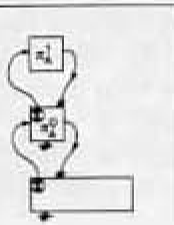
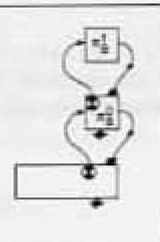
However, we deliberately avoided reference to the preordination of C^+, C^- and C^X because, in this respect, his theory and our theory are addressed to entirely different problems (as mentioned, very briefly, in Section 3.2).

Player a and Player b may be (but are not necessarily) the participants A and B. If they are (so that a is A and b is B), if the joint modelling or game like situation is named "i" on occasion n, then <u>Strat</u> a = <u>Strat</u> A, <u>Strat</u> b = <u>Strat</u> B, and these are serially executable models constructed by $\underline{\text{Proc}}^0_A$ i, $\underline{\text{Proc}}^0_B$ i; on an occasion n. If execution takes place and terminates then $c \in C^{\infty} \subset C$ because a certain relation, tentatively entitled r_i, holds between the execution sequences in the context of the modelling facility (plausibly, but extratheoretically, $c \in C^+$ if $r_i = R_i$ and R_i is satisfied).

At this level, Howard's theory corresponds very closely to our theory of conversations. The preference graphs ψ (for Player 1 identified with participant A and Player 2 identified with Participant B, the graphs ψ_A, ψ_B) reflect the intentions of these participants in a context that is labelled "i". That is, to say that "A has a preference function ψ_A and B a preference function ψ_B" (it was noted earlier that these serve to characterise A and B) is just to say that "on occasion n the L^0 interactions of A and B (if they exist) are governed by $\underline{\text{Proc}}^0_A$ i and $\underline{\text{Proc}}^0_B$ i". These are concepts, procedure classes, or (in a non-trivial sense) goals that lead A and B to build models <u>Strat</u> A and <u>Strat</u> B that, on execution in the modelling facility, give rise to an outcome in the "basic game" (invariably, in our examples, a partially cooperative game). But that is not the end of the matter.

4.6. The strict conversation between A and B on occasion n is the microdynamics of a series of metagames over the basic game. To convert the microdynamic or cognitive or computational picture into terms of the metagames over the basic game it is necessary to pass (as we did in Chapter 5 and Chapter 6) from *executions* or *explanations* ($\underline{\text{Exec}}^0$ i, $\underline{\text{Exec}}^1$ i, or $\underline{\text{Expl}}^0$ i, $\underline{\text{Expl}}^1$ i) to <u>PQuests</u> (or the alternative sets over which they are posed) and (insofar as B questions or commands A) to <u>Req</u> i and/or <u>Enq</u> i. The notation needed for this purpose is furnished by Table 1, and Table 2 of Chapter 6; we shall also assume (in line with the earlier discussion of uncertainty regulation) that A and B reply by confidence estimates $\underline{\text{Eval}}^0_A$ i to $\underline{\text{PQuest}}^0_A$ i, by $\underline{\text{Eval}}^1_A$ i to $\underline{\text{PQuest}}^1_A$ i; by $\underline{\text{Eval}}^0_B$ i to either $\underline{\text{Req}}^0$ i or $\underline{\text{Enq}}^0$ i, and by $\underline{\text{Eval}}^1_B$ i to either $\underline{\text{Req}}^1$ i or $\underline{\text{Enq}}^1$ i.

Fig. 9.9. Processes underlying the metagame of a strict conversation. All transactions refer to alternatives, and choices amongst alternatives (here, elicited as confidence estimates) are determined if the strict conversation is represented as a metagame or a series of metagames.

Conversational processes	Metagames	Transactions in a strict conversation (or internal interactions) and corresponding alternative sets
	g	$\langle \underline{Strat_A} i, \underline{Strat_B} i \rangle$
	A, g	$\langle \underline{PQuest_A} i, \underline{Expl^0} i \rangle$
	B, g	$\langle \underline{Req^0_B} i, \underline{Expl_A} i \rangle$
	A, B, g	$\langle \underline{PQuest^1_A} i, \underline{Expl^1_A} i \rangle$
	B, A, g	$\langle \underline{Req^1_B} i, \underline{Expl^1_B} i \rangle$
	A, A, g	Internal
	B, B, g	Internal

The dynamic processes which correspond to the various metagames (the actual or internal-and-cognitive processes of reflection to which Howard alludes) are dissected out in Fig. 9. These processes (with explanations or executions truncated to form alternative sets) are metagames, rather than the processes that underlie them. The original notation (appropriate in the context of Chapter 6) is retained for the sake of consistency though it is otherwise perverse to write $\underline{Req}^0$i instead of $\underline{PQuest}^0_B$i and $\underline{Req}^1$i instead of $\underline{PQuest}^1_B$i.

To say "A and B agree, on occasion n in a strict conversation, that a commonly named relation R_i is brought about", is equivalent, in this picture, to saying "The solution of i is general metarational for A and B" and (if, as usual, the strict conversation is tutorial; when B demands that the extension of A's model is R_i) that R_i is somehow satisfied.

The strict conversation (if occasion n is terminated by agreement) depicts the mechanism of the metagame played by participants A and B on occasion n; notice that Howard's "players" are, as they should be, P Individuals.

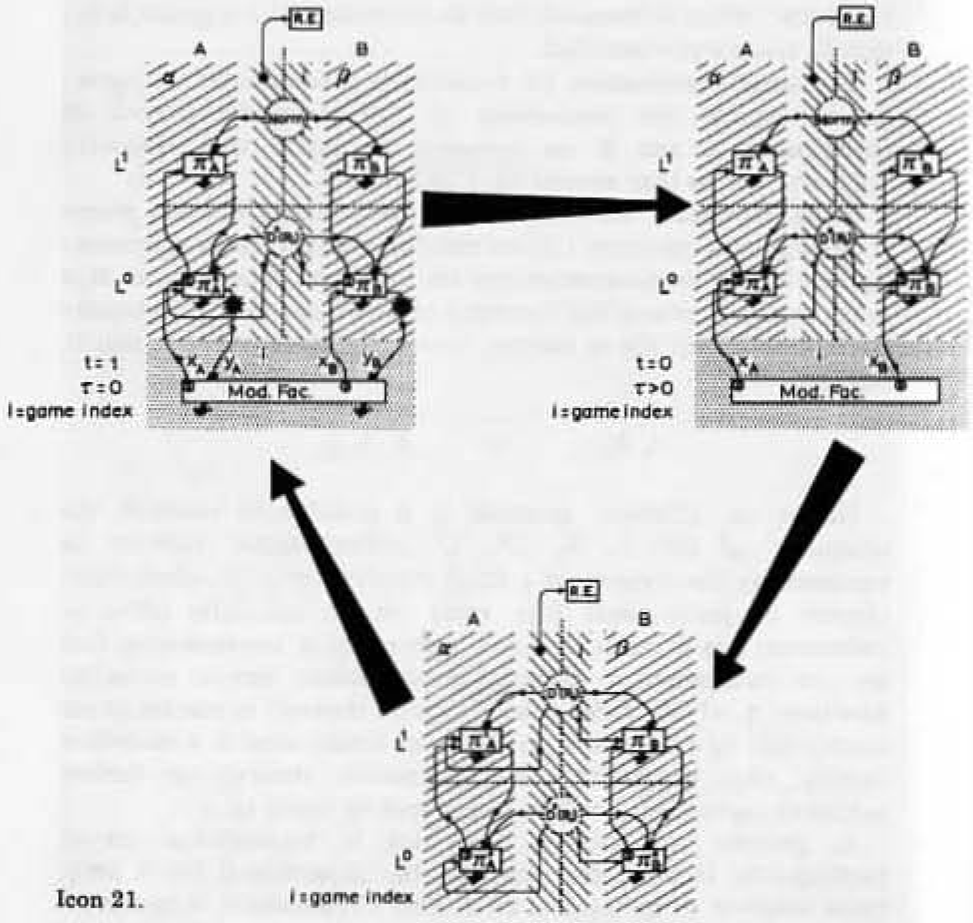
From Howard's Theorem 1, (in particular, from the Prime Representative corollary 1.2) we may infer that so far as agreement (in our sense) or metarationality (in his sense) is concerned, it is unnecessary to extend the hierarchy of reflexions to any metagame beyond order m! for m players; hence, for participants A and B, beyond

$$A, B, g \qquad \text{or} \qquad B, A, g.$$

Hence, on different grounds, it is possible to establish the adequacy of the A, B, L^0, L^1 conversational skeleton as representing the content of a strict conversation. Of course either players or participants may carry on an indefinite spiral of reflexions; this is conceeded and is common in conversations that are not constrained to be strict conversations. But, so far as the functions η, η^* are defensible (Howard's theory), or insofar as the satisfaction of a relation R_i is tied to or instantiated in a modelling facility (the jargon of our conversation theory) no further solutions (agreements) will be uncovered by doing so.

In general, an outcome need not be metarational for all participants. It may, for example, be metarational for A only, from some or all metagames of g. This circumstance is important in the next sections where issues such as "A disagreeing with B over R_i but knowing why B has a different point of view" receive attention. It simply happens that in a strict tutorial conversation (the case most thoroughly discussed) any occasion n is ended by an A, B agreement, since that is one condition for understanding.

Howard specifies players and coalitions in a uniform fashion. So, for example, the A, B, conversation is a possible coalition. Our conversation theory maintains a similar uniformity at the microstructural level. The A, B conversation is, as it should be, a P Individual like a participant. Starting from this comment and mustering far deeper aspects of metagame theory than those sketched in this section, it is possible to develop a conversational

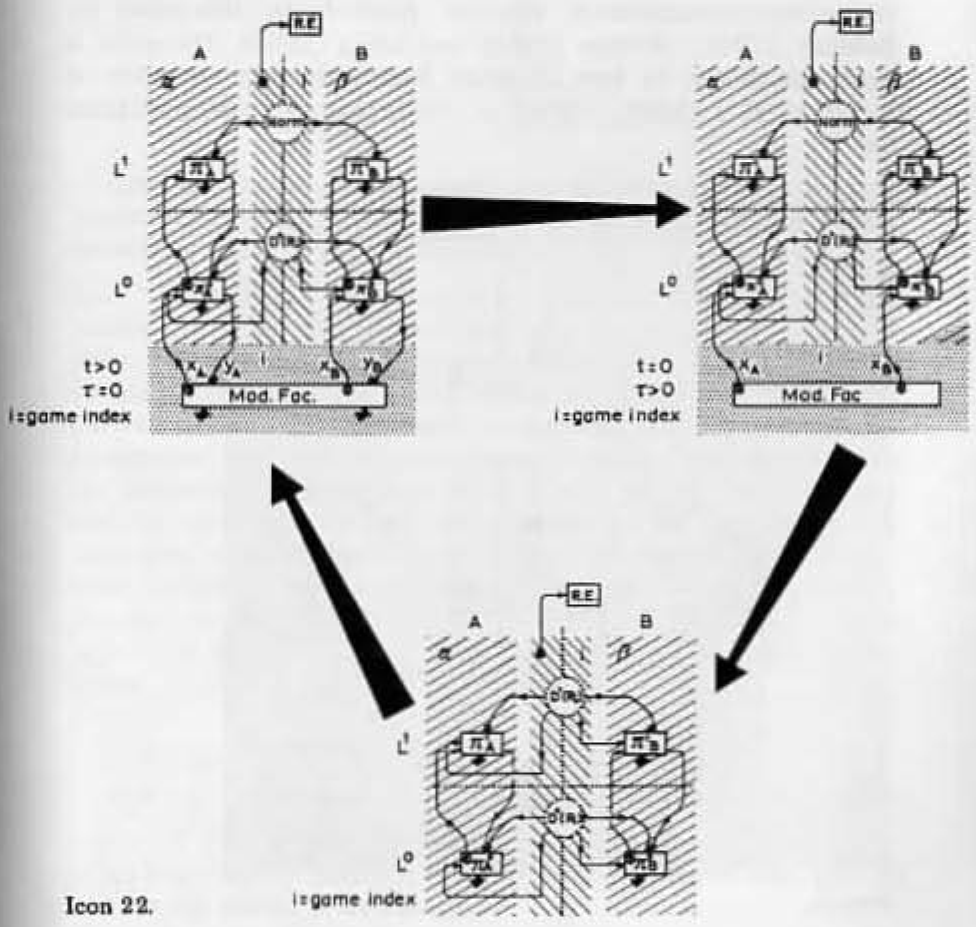


Icon 21.

theory of social and political interaction. Some initial essays in this direction will be reported in the next volume.

Here, we condense the computational pictograms of Fig. 9 in terms of the standard iconic notation, for an A, B, system only.

4.7. Icon 21 (the play of a metagame, in abbreviated but exact form) may thus be compared with Icon 20 (the play of a standard game). In the standard game A and B are to contemplate a fixed norm; in the metagame A and B entertain metahypotheses that act as norms. The L^0 and L^1 reflections may involve real life events (Howard's bargainings and questionings) or cognitive events (hypothesis formation).

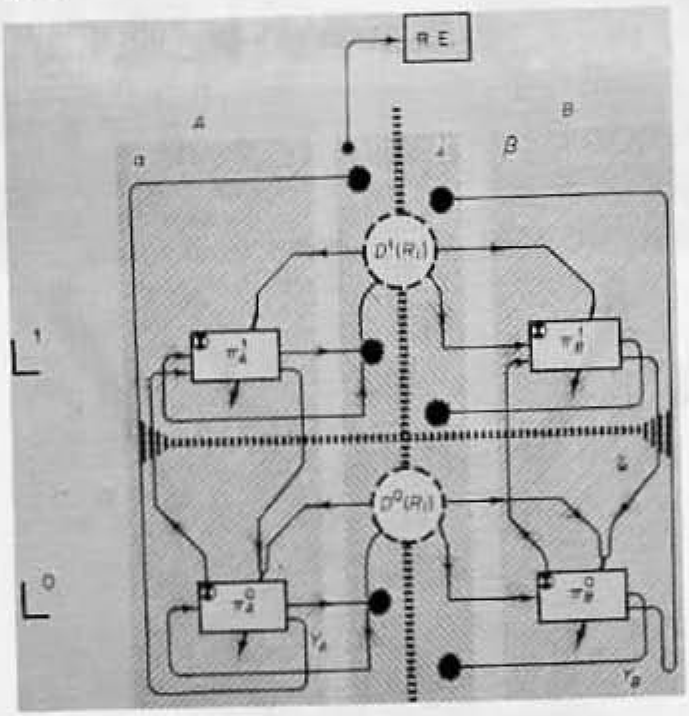


Icon 22.

We can pass from the (selective) case (2) of Section 3.4 to the constructive case (1) of Section 3.4 by means of Icon 22 (which extends Icon 21 in the same way that Icon 20 extends Icon 19). But the picture is vacuous insofar as the entity corresponding to the joint or composite automaton of Icon 22 may not, by previous definition, be executed in a one clocked modelling facility. This entity is, in fact, A's image of B and B's image of A.

5. An interlude: interpersonal test measures

The selective metagame construction (Icon 21) images a cluster of important test and sampling techniques, which empirically underpin the whole of transactionalist psychology and those normative/communication oriented psychologics developed by Bateson, (1956), Brodey (1968) and Laing (1972). The point is made specifically by Icon 23 which depicts the administration of Laing et al.'s (1966) "IPM" or "interpersonal communication test".



Icon 23.

5.1. Let A and B represent two parties (marital partners or other mutually dependent individuals). On the assumption that A and B are usually in regular conversation, they are isolated and presented with a series of suitably pronominalised questions about relations, R_i, that they may maintain or fail to maintain (for example, in psychotherapy, R_i may be "loves me"). Each question in the battery has the form:

A? comment on R_i $= A(R_i)$
A? What is B's comment on R_i $= A(B(R_i))$
A? What is your comment on B's comment on R_i
$= A(B(A(R_i)))$

Similarly, for B, the question set

$B(R_i)$
$B(A(R_i))$
$B(A(B(R_i)))$

The undetermined *comment on* is substituted, for each occurrence, by a particular comment, such as a rating of the validity of a statement that R_i holds.

5.2. There is no objection to using the same technique in the context of less emotively loaded relations and in this case it is often convenient to elicit self-referential IPM responses by changing the respondent's situation or viewpoint. For example, in our own work on driving strategies (Pask et al. 1971), we present one respondent with various series of photographs. Each series shows the respondent a drivers eye view (from "his" vehicle) of another vehicle, and leads up to a potentially hazardous interaction occurring past the "end" of the series. The respondent gives an IPM response as a degree of belief (confidence estimate) over the possible outcomes of the interaction, the outcome depending in part upon his own action, and in part upon the other drivers action; thus

$A(R_i)$ your belief that you (A) will swerve, or decelerate.

$A(B(R_i))$ your (A's) belief that the other driver (B) will swerve or decelerate.

$A(B(A(R_i)))$ Your (A's) belief that B believes that you (A) will swerve or decelerate.

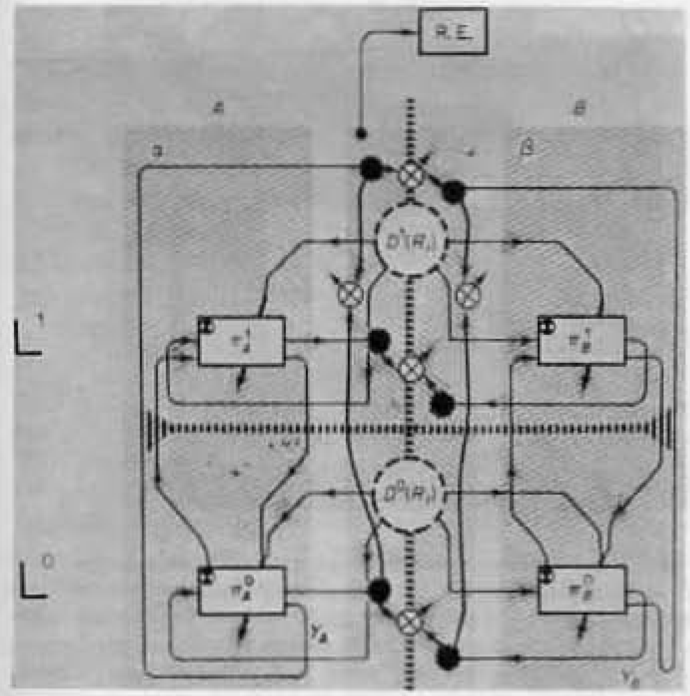
Each photographic series has a dual, depicting the same situation from the other driver's point of view. Consequently, when the same respondent is required to express his IPM degrees of belief in respect of the dual of situation R_i, he is placed in the other driver's (B's) role and his responses become

$$B(R_i)$$
$$B(A(R_i))$$
$$B(A(B(R_i))).$$

5.3. In any application of the IPM the hierarchical responses are matched to obtain indices of "agreement" (comparison of $A(R_i)$ and $B(R_i)$) of "mutual comprehension" (comparison of $A(B(R_i))$ and $B(A(R_i))$) and other items such as self reflexivity (comparison of $A(B(A(R_i)))$ and $A(R_i)$ or $B(A(B(R_i)))$ and $B(R_i)$). In the case of Section 5.1 the numbers are derived from two respondents; in the case of Section 5.2 from one respondent in two different roles. As might be anticipated, the indices vary, to a large extent, independently. For example, agreement does not necessarily imply mutual comprehension, or vice versa. In the driving strategy study, for instance, it was obvious that either misperception of R_i or misperception of the other driver's perception of R_i can lead drivers to entertain beliefs likely to engender hazardous behaviour. It is also true that agreement ($A(R_i)$ compared with $B(R_i)$) is often independent of the mutual comprehension index obtained by comparing $A(B(R_i))$ with $B(A(R_i))$.

5.4. In Icon 23 the (possibly independent) hierarchical test data from A and B are the (selective) strategies of players A and B in a metagame. The players do not interact directly (the test batteries are separately administered) and the data is accumulated by an experimenter, throughout "play". For situations such as the driving study of Section 5.2 the metagame is solitaire. At the end of the playing phase, the players are shown an outcome which may either be performance based (for example, in driving, an index of safety for each distinct dual series of photographs) or it may consist in a profile (of "agreement", of "mutual comprehension", and so on).

5.5. Alternatively, the profile indices may be fed back as a near continuous corrective signal, as in Icon 24. In practice, it is essential to mechanise the calculation. We use the equipment,



Icon 24.

FRIM (Feedback Regulated Interpersonal Method), of Fig. 10 for this purpose.

5.6. The hierarchy of indices could, in principle, be continued indefinitely [for example, to obtain $A(B(A(B(A(R_i)))))$]. In fact, it is truncated at a level labelled L^2. The rationale of truncation is usually stated along quite commonsensical lines by noting that the indices loop back upon each other at L^2; thus $A(B(A(R_i)))$ denoted the same thing as $A(R_i)$ and vice versa for B. The argument is sound enough, provided that the meaning of R_i is unchanged by change of level, (for example, in driving, that the consequences of deviating from a safe "between vehicle", R_i, is an accident involving both drivers, regardless of how the blame might be assigned). In general, this condition is satisfied, also, by the effectively interpersonal systems for which the IPM test is used in psychiatry.

It is some interest that truncation is justified on the grounds

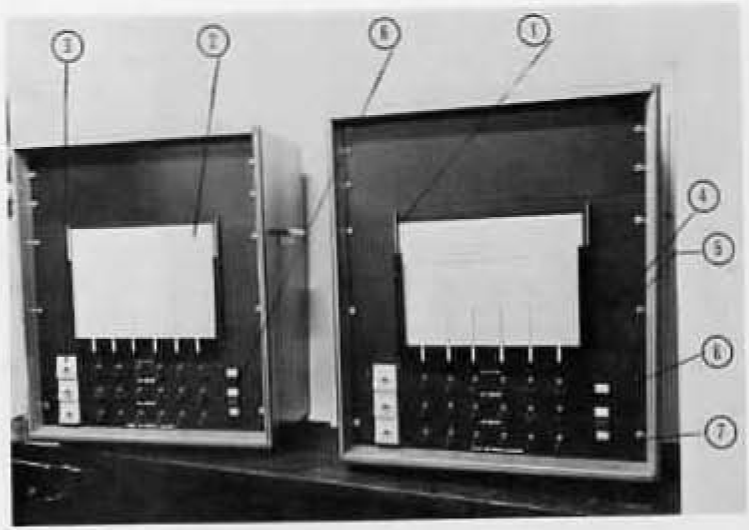


Fig. 9.10. Two identical FRIM consoles one for each two participants A and B who interact as shown in Icon 24. Questions centred upon a topic relation R_i are inscribed on cards (1) together with punched-hole designating-codes. The cards are inserted in photoelectric card readers (2), the codes indicate (to CASTE or other operating system) topic and response type; namely, confidence estimate (meters (3) as for BOSS), multiple choice, or assignment of values +, − and null (or irrelevant) to questions items. (3) positions switch banks (4) with register lamps (5), + = on, − = flashing, null = off). Participant A enters responses $A(R_i)$, $A(B(R_i))$ and $A(B(A(R_i)))$ on upper, middle and lower response switch banks; B independently enters $B(R_i)$, $B(A(R_i))$ and $B(A(B(R_i)))$. Response process is sequenced by indicator lamps (6). After that, feedback lamps (7) display $A(B(A(R_i)))$ to B, and $B(A(B(R_i)))$ to A; other responses may be modified to reach compromise, in light of mutual feedback. Next, feedback lamps (8) display $A(B(R_i))$ to B and $B(A(R_i))$ to B; the participants may alter $A(R_i)$ and $B(R_i)$. Finally, A receives feedback of final $B(R_i)$ and A receives feedback of final $A(R_i)$ and question cards are withdrawn. Each stage in the process is recorded and transmitted to the operating system.

that further expansions of the metagame (adding further levels to the hierarchy), would uncover no essentially novel solutions (or, in this case, no further stable measure points).

In turn, this is a special case of the truncation requirement of conversation theory (Chapter 5 and Chapter 6) which is used to demarcate the minimal conversational skeleton.

6. Modelling Facilities for tasks involving interpersonal interaction

Certain educationally significant tasks call for the art of characterisation; for example, a history student is required to model (quite literally, to write a procedure called the character of) an historical person, to fathom this person's motives, intentions, and so on; indeed, to enact the person's role or (rephrasing it) to execute the procedures he has written or modelled. Similar feats of characterisation are called for in literary studies and a whole gamut of less specialised creative occupations.

With one exception (mooted below) it has not been possible to realise such acts as part of a tutorial conversation. But the matter of doing so is approached, theoretically, as an extension of the (selective metagame) Icon 21 to its constructive analogue (Icon 22). As noted at the end of Section 4 the entity to be constructed (a procedure called "the character") cannot be executed in a modelling facility, of the sort we have at hand, with the constraints of Fig. 1 in Chapter 5. There is no serious obstacle on the path to providing a facility as liberal as Fig. 2 of Chapter 5 and one of them is currently under construction. By way of conjecture, it is possible to build and (concurrently) to execute characterising procedures in a modelling facility of this type.

Some attention has been given to one special, ubiquitous, but often overlooked case of characterisation; namely, the case of teaching. Thus a genuine tutor, in contrast to an uncertainty regulation heuristic, or the like, is capable of modelling (in the sense of the last paragraph of "characterising") the student; vice versa, the student models the tutor. Moreover, the execution of the tutor's (student) model, perhaps under specific limiting conditions, is a prerequisite for non trivial teaching. An approximation to that arrangement, using the FRIM system mentioned in Section 5.6 is operational. This, and proposals for a complete realisation, are discussed in the sequel.